University of Milano-Bicocca

Doctoral Program in Experimental Psychology, Linguistics and Cognitive Neuroscience.

34° Ciclo

2015 – 2018

Milano

Visual Illusions and Fourier analysis as psychophysical tools to support the existence of the Number Sense

Andrea Adriano

Doctoral Thesis promoted by:

Prof. Luisa Girelli

“La Natura è un libro scritto in caratteri matematici”

Galileo Galilei

Table of Contents

Abstract

The natural environment in which animals are forced to survive shapes their brain and the way in which they behave to adapt and overcome natural pressures. These selective pressures may have determined the emergence of an evolutionary ancient neural system suited to rapidly extract abstract information from collections, such as their numerosity, to take informed decisions pivotal for survivance and adaptation. The "Number Sense" theory represents the most influential neural model accounting for neuropsychological and psychophysical evidence in humans and animals. However, this model is still largely debated because of the methodological difficulties in isolating neural signals related to "discrete" (i.e., the real number of objects in a collection) abstract numerosity processing from those related to other features correlated or confounded with numerosity in the raw sensory input (e.g., visual area, density, spatial frequency, etc.). The present thesis aimed to investigate which mechanisms might be at the basis of *visual* numerosity representations, overcoming the difficulties in isolating discrete from continuous features. After reviewing the main theoretical models and findings from the literature (**Chapter 1 and 2**), in the **Chapter 3** we presented a psychophysical paradigm in which Kanizsa-like illusory contours (ICs) lines were used to manipulate the connectedness (e.g., grouping strength) of the items in the set, controlling all the continuous features across connectedness levels. We showed that numerosity was underestimated when connections increased, suggesting that numerosity relies on segmented perceptual objects rather than on raw low-level features. In **Chapter 4**, we controlled for illusory brightness confounds accompanying ICs. Exploiting perceptual properties of the *reverse-contrast* Kanizsa illusion, we found that underestimation was insensitive to inducer contrast direction, suggesting that the effect was specifically induced by a sign invariant boundary grouping and not due to perceived brightness confounds. In **Chapter 5**, we concurrently manipulated grouping with ICs lines and the perceived size of the collections using classic size illusions (Ponzo Illusion). By using a combination of visual illusions, we showed that numerosity perception is not based on perceived continuous cues, despite continuous cue might affect numerical perception. In **Chapter 6** we tackled the issue with a direct physical approach: using Fourier analysis to equalize spatial frequency (SF) in the stimuli, we showed that stimulus energy is not involved in numerosity representation. Rather segmentation of the items and perceptual organization explained our main findings. In **Chapter 7** we also showed that the ratio effect, an important hallmark of Weber-like encoding of numerical perception, is not primarily explained by stimulus energy or SF. Finally, in

Chapter 8, we also provided the first empirical evidence that non-symbolic numerosity are represented spatially regardless of the physical SF content of the stimuli. Overall, this thesis strongly supports the view that numerosity processing is not merely based on low-level features, and rather clearly suggests that discrete information is at the core of the Number Sense.

Acknowledgments

This work originated from a fruitful discussion that I had with Dr. Luca Rinaldi at the beginning of the PhD in Milan. I must admit he was the first person believing on the successful realization of this project. For this reason, I would like to give him a special thank over and above all the others. Luca helped me in developing my experiments and he carefully read and commented all my drafts. He improved the scientific quality of all the things with his commentaries and useful feedbacks. Thank you, Luca, for your precious help and for all the good moments that we spent on the pitch! You are a lethal striker and a great captain/coach too!

Equally important, has been also the role of my beloved advisor, Prof. Luisa Girelli, who helped me in developing my personal ideas in a comfortable and stimulating academic context. She allowed me to work on a challenging topic giving me the support for everything, including strong moral and psychological support, during the PhD period and the pandemic isolation. Unfortunately, the pandemic forced us to work from home and to meet on Skype, but I always appreciated your "presence". I was honored to work under your kindly supervision. As PhD student, I hope that my work satisfied your expectations.

I cannot forget Dr. Lola De Hevia, who hosted me in her fantastic baby-lab at the university of Paris-Descartes during the last 4 months of my PhD. She is an incredibly smart and charismatic researcher with a contagious "Mediterranean" good mood! But above all, she is a very humble person. Thank you for all, Lola!

I want to also thank my old "southern" friends living in Milan: Gianluca, Mariangela, Federica, Raul, Giusy, and Francesca. Unfortunately, the work will lead me another time far away from you, but we live in the same place in the South Italy, so we will meet again and again as usual near to our beautiful sea. And thank you to all the fantastic Professors and my colleagues of Milano-Bicocca. I will never forget the fantastic days in the University of Milano-Bicocca, before the pandemic! A special thank also to my Italian colleagues at the University of Paris-Descartes, Dr. Gisella Decarli and Dr. Ludovica Veggiotti, and last but not least, Dr. Martina Corazzol who supported me a lot.

Finally, thank a lot to my family for giving me all the support and love during the last 3 years. This gives me the force to overcome any trouble during this long journey. But life is a journey, yet full of difficulties, but *"Je n'ai pas peur de la route"…*

List of Figures

Chapter 1

1.1. General Introduction

In everyday life, we are often faced with practical "mathematical" problems that can be rapidly and successfully solved using raw approximate numerical information to take optimal decisions. For example, we can quickly extract "at a glance" the approximate number of people in a crowded student room, or we can rapidly choose the line containing the fewest people at the post office. This apparently effortless ability to extract numerical information, without counting the items one-by-one, has been scientifically investigated in a remarkable article published in *Nature* almost 150 years ago by the British economist W.S. Jevons, who empirically showed that people can rapidly report the approximate number of beans threw into a box with an extreme precision for set-size of around 3 or 4 beans, and with error in estimations that increased with set size for larger amounts (Jevons, 1871). Nowadays, according to a vast amount of multidisciplinary research, spanning from animal cognition to psychophysics and cognitive neuroscience, this intuitive elaboration of numerosity (generally referred to as "Number sense"), is thought to be supported by dedicated neural networks in the brain suited for the processing of both small and large numerical quantities. In particular, for smaller numerosities (e.g., 3-4 items) it is assumed that numerical information is processed by a very precise system (*Subitizing*), while for larger amounts another system would be at play: the so-called Approximate Number System (ANS; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). Although the two systems might be supported by different cognitive mechanisms and are defined by distinct behavioral signatures, in this thesis we will mainly focus over the range of the ANS (numerosities larger than 4), also because its total independence from the subitizing is still a matter of intense debate (e.g., Cordes, Gelman, Gallistel, & Whalen 2001).

In the scientific literature, the ANS is defined as an evolutionary ancient neural network shared across different animal species (e.g., Agrillo, Dadda, & Bisazza, 2007; Agrillo, Dadda, Serena, & Bisazza, 2008; Agrillo, Dadda, Serena, & Bisazza, 2009; Agrillo, Piffer, & Bisazza, 2011; Agrillo, Piffer, Bisazza, & Butterworth, 2012; Brannon & Terrace, 1998; Ditz & Nieder, 2015), and already active in early months of life even in children (e.g., Brannon, Abbott, & Lutz, 2004; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). This system would allow living beings to extract approximate numerical information from sensory input stimuli: as such, the ANS may have originally evolved in

the animal kingdom as an adaptive mechanism to solve numerical tasks necessary for survival in natural environment (e.g., Nieder, 2021). From an evolutionary standpoint, this testifies the importance of a dedicated neurocognitive system for solving numerical problems that are, in turn, essential for survival in the physical environment. Among many other requests, indeed, animals may need to compare the number of conspecifics, the number of rivals, or they might select the larger set when looking for food (e.g., Agrillo et al., 2012; Benson-Amram, Heinen, Dryer, & Holekamp, 2011; Perdue, Talbot, Stone, & Beran, 2012; Piantadosi & Cantlon, 2017).

The primary behavioral signature of the ANS is the *Weber's law* compliance, which holds universally in different human cultures, as well as across human development and different species (e.g., Whalen, Gallistel, & Gelman, 1999). Weber's law states that the threshold of discrimination between two stimuli increases linearly with stimulus intensity or, accordingly to its strong formulation, that the probability to correctly discriminate two stimuli depends on the ratio of their intensities (Brus, Heng, & Polanía, 2019; Dehaene, 2003). The compliance with Weber's law has been mainly investigated in numerical comparison tasks, whereby participants have to choose the numerically larger (or smaller) set between two rapidly presented arrays of items. In these tasks, error rates and reaction times typically increase as a function of the numerical ratio (i.e., smaller numerosity/larger numerosity) between the two arrays to be compared. Such ratio-dependent behavior has been taken as a signature that non-symbolic number processing follows Weber's law (e.g., Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). Importantly, the Weber Fraction (i.e., indexing the precision of numerical representations) has been shown to predict later symbolic mathematical achievements in children and it is typically lower in dyscalculic populations compared to controls (Halberda, Mazzocco, & Feigenson, 2008; Piazza, Facoetti, Trussardi, Berteletti, Conte, Lucangeli, Dehaene, & Zorzi, 2010). From a neural standpoint, neurons tuned to a preferred numerosity have been found in human and macaque parietal cortex (e.g., Castelli, Glaser, & Butterworth, 2006; Harvey, Klein, Petridou, & Dumoulin, 2013; Nieder & Miller, 2004; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004).

Yet, despite the overall consensus about the existence of the ANS, the exact visual mechanisms underlying this system are still a matter of intense debate, with several models that have been advanced to illustrate how numerosity is possibly extracted from visual input (e.g., Dakin, Tibber, Greenwood, & Morgan, 2011; Dehaene & Changeux, 1993; Meck & Church, 1983). On the one hand, mainstream psychophysical and computational models hold that the ANS is based on a primitive visual segmentation and individuation algorithm. According to these models, numerosity

would be directly extracted on individual objects independently from their physical features like shape, size, position, etc. (Burr & Ross, 2008; Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012; Verguts & Fias, 2004). Dehaene and Changeux (1993) first proposed an influential computational model based on spatially segregated items and composed of two main computational blocks: the first block would filter non-numerical features from the visual input ("normalization" stage), whereas the second block would classify the normalized activity and provide it with a numerical label ("classification" stage).

However, the idea that numerosity perception acts as a primary sense has been challenged by recent accounts. Recent theories maintain that numerosity representation would be indirectly extracted exploiting one or more surrogate features correlated or confounded with the numerosity in the set (Allik & Tuulmets, 1991; Dakin et al., 2011; Durgin, 2008; Gebuis & Reynvoet, 2012a, 2012b, 2012c): the so-called *continuous variables* (e.g., total surface occupied by the set, density, luminance, etc.). The view that features confounded with numerosity may bias numerical perception is certainly not new and dates back to the early 90s, with the *occupancy model* proposed by Allik and Tuulmets (1991). According to this model, numerosity is estimated by summating the area occupied by the items in the set, assuming that each item occupies a "virtual" area defined by an occupancy radius, rather than its physical extent. The more the virtual areas occupied by two neighboring items overlap, the more the numerosity would be underestimated. In line with this prediction, findings from behavioral tasks indicate that perceived numerosity decreases as the spacing between the elements is reduced (Ginsburg & Goldstein, 1987).

Besides the impact of these variables, several other continuous features have been shown to affect numerosity discrimination, including the item size, the aggregate surface, the extent and shape of the convex hull (e.g., the virtual elastic enclosing the items), and the density (e.g., Gebuis & Reynvoet, 2012a; Hurewitz, Gelman, & Schnitzer, 2006; Katzin, Katzin, Rosén, Henik, & Salti, 2020). In brief, since it seems physically impossible to have two different numerosities with the same physical amount of non-numerical features, it has been suggested that numerosity would be indirectly extracted by integrating one or more surrogate continuous variables co-varying in the natural world with numerosity (Gebuis, Kadosh, & Gevers, 2016; Leibovich, Katzin, Harel, & Henik, 2017), without the need to suppose the existence of a neural system dedicated to extract a discrete representation of quantity. In addition, some authors advocate that the development of a dedicated numerical system exploiting discrete "numerosity" would be "expensive" for the brain (Gebuis & Reynvoet, 2011). Indeed, since numerosity is confounded with continuous cues such as area and

density, and with other particular texture statistics (e.g., spatial frequency), the brain should easily exploit these visual features to represent numbers. Although this argument may appear biologically plausible, it should be remembered that sometimes the cost for developing a neural structure could be compensated by a new evolutionary advantage.

This thesis aimed to show that discrete numerosity and low-level *continuous features* are two different level of analysis of the visual input. In other words, we aim to show that a neural system that extract *only* low-level continuous features or texture statistics (as suggested by alternative models of Number Sense) would fail to explain several findings in the literature of the ANS and would lead the observers to non-optimal choices. By using visual illusions and Fourier analysis, in this thesis we will demonstrate that Numerical perception is not based on continuous cues. In particular, we will suggest that Perceptual Organization may play an important role in the structuring of the scene, shaping how we perceive a coherent *discrete* set of perceptual objects (numerosity).

In **Chapter 1**, we will briefly present the available evidence in the animal, children and human literature and the typical behavioral and neural signatures of the ANS. In **Chapter 2**, we will introduce more in details the problem of continuous cues and the alternative models on Numerical perception proposed in the literature. We will also bring evidence from studies linking numerical perception to Gestalt psychology and visual illusions. In the **Chapters 3 to 8** we will then move on by presenting 6 different experimental works that used Visual Illusions and Fourier Analysis to disentangle numerosity from low level sensory continuous cues.

1.2. Numerical abilities without language: An innate number sense?

1.2.1. Animal Research

As we briefly introduced in the previous paragraph, a large amount of work in animal cognition and comparative psychology suggests that animals may have non-verbal numerical discrimination abilities, intermodal processing of numerosity and elementary arithmetic abilities comparable to those of humans. In animals, these skills could be crucial for solving various problems that arise in their natural environment pivotal for survival and adaptation. Indeed, numerical judgments in animals are very important because they may allow the elaboration of anti-predator strategies, as the probability of being captured decreases when individual animals join a large group of conspecifics (e.g., Agrillo et al., 2012), or because they allow to optimize the food intake, quickly choosing the largest quantity between two sources of nutrition (e.g., Beran, Evans, Leighty, Harris, & Rice, 2008). Furthermore, these numerical skills are also useful in social interactions, as some species of animals tend to attack other groups of conspecifics more when they perceive themselves as part of a larger group (Wilson, Britton, & Franks, 2002). Therefore, various animal species, such as rats, pigeons, parrots, fish, monkeys, chimpanzees and others, possess a sense of number.

One significant evidence of numerical abilities in animals was reported by Meck and Church (1983), who trained mice to press one lever in response to a two-tone sequence and to press another lever in response to an eight-tone sequence. They observed that the discrimination of the duration of the sounds was good in the initial phase and that subsequently the mice generalized their behavior to new sequences, whose duration was fixed and varied only in the number of tones: for the same number of tones, they discriminated sequences of two seconds from sequences of eight, while for sequences of equal duration the mice discriminated two tones from eight. This suggested that the animals had formed an abstract representation of quantity during the initial training phase and, thanks to it, they were able to discriminate the stimuli. In a later experiment (1984) they conditioned mice to press the lever to the left of the cage when they heard two-tone sequences and to press the right lever when they heard four tones. The animals were then separately trained to associate two flashes with the left lever and four flashes with the right lever. The researchers then presented combined auditory and visual stimuli and observed that if a synchronous tone and flash were presented, i.e., a total of two events, the mice pressed the left lever, while if two tones and two flashes were presented, they pressed the one on the right. The

results thus showed that the mice generalized their initial learning to new conditions and that, consequently, their behavior was based on the abstract concept of number, that did not depend on the specific sensory modality in which input stimuli occurs.

Further research investigated the ability to order and discriminate also quantities in other animal species. For example, Brannon and Terrace (1998) conducted an experiment with rhesus monkeys (in a setting in which non-numerical variables of stimuli were controlled). Initially monkeys underwent a training in which they learned to sort in ascending order panels that contained a number of elements between 1 and 4. Later, their ability to order pairs was tested with a number of elements between 5 and 9 (Figure 1.1). The results showed that monkeys were able to order stimuli in ascending order: that is, they created a representation of the new numerosities, applying the same ordering rule to the new stimuli. Subsequently, Cantlon and Brannon (2006) proved that the same family of monkeys can extend this rule of increasing order to larger numbers (10, 20 and 30). In addition, Cantlon and Brannon's (2007) work showed that monkeys spontaneously create number representations and use them to discriminate between different quantities.

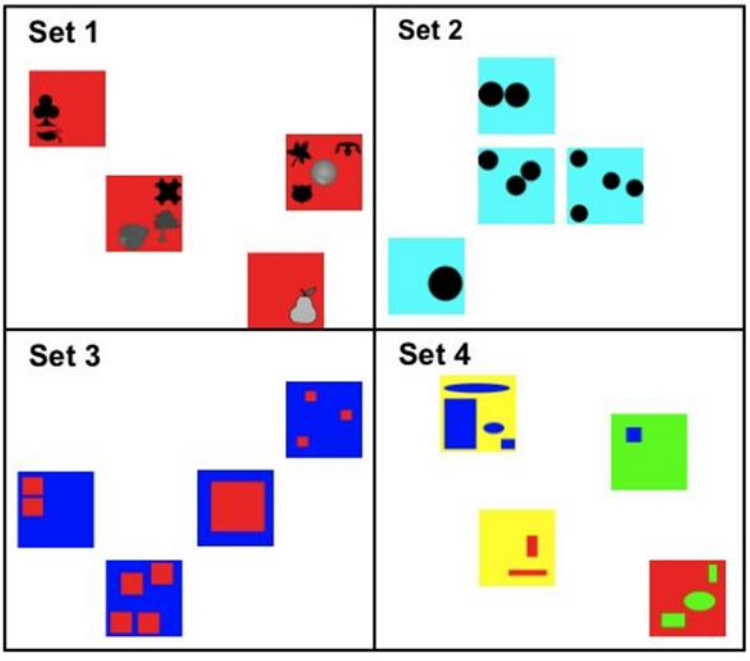


Figure 1.1: Example of stimuli used to train monkeys in the study of Brannon and Terrace (1998).

To fully understand the evolutionary path of the Number Sense, research was also conducted on birds and fish, investigating the brain structures involved in number processing and comparing them with those of mammals, in particular primates. In an electrophysiological study conducted by Ditz and Nieder (2015) on crows, it emerged that, despite the enormous difference in the

neuroanatomy of birds, the neurons in an area called nidopallium caudolateral (NCL) were tuned to a particular number (Figure 1.2).

Figure 1.2: Example of stimuli used to train crows in the study of Ditz and Nieder (2015).

Furthermore, the study by Miletto Petrazzini and Agrillo (2016) investigated numerical abilities in a species of fish (*Poecilia reticulata*), involving adult females. The choice of exclusively female subjects was justified by evidence of the high sociality of females of this species (Miletto Petrazzini & Agrillo, 2016), which naturally leads them to form groups of different sizes and to join large groups already formed.

The research design involved testing the fish under two experimental conditions. In one condition the fish was placed in the center of a thank and two groups of peers with a set size within the subitizing range (2 vs 3 individuals) were placed on the two lateral sides, while in the other the group size was outside the subitizing range (6 vs 10 individuals). The results showed that fish spent more time in both conditions (e.g., in-range condition and out-of-range condition) in closer proximity to the largest group of peers (e.g., 3 or 10). While these results leave open several questions about the presence of two distinct number systems in these animals (e.g., Subitizing system and ANS), on the other hand they demonstrate how these animals are able to discriminate in a non-random way between two sets that differ in terms of quantity of elements. This suggests a common phylogenetically ancient origin of non-verbal numerical systems in all vertebrates (for a review Agrillo & Bisazza, 2018).

1.2.2. Children Research

The investigation of children's numerical competences began around the '80s, as before developmental psychology was mainly dominated by the constructivist perspective, which argued that mathematical and logical abilities were progressively built in the child's mind through observation, internalization and abstraction of the regularities of the external environment. The famous developmental psychologist Jean Piaget, in fact, believed that at birth children lacked arithmetic skills and that they were unable to learn and apply them before the age of 4-7 (Piaget, 1952). On the other hand, several experimental studies have underlined the limits of this vision and have highlighted the presence of the non-symbolic numerical systems and the ability to perform simple operations even in children of preverbal age (Starkey & Cooper, 1980; Simon, Hespos, & Rochat, 1995; Wynn, 1992), suggesting an ontogenetically pre-determined development of these abilities in humans.

Starkey and Cooper (1980) were the first to demonstrate the presence of a numerical discrimination ability in 6-7-months-old children using the visual habituation paradigm. In the first experimental condition (e.g., habituation phase), children were repeatedly shown a series of slides with a fixed number of 2 points until the fixation time decreased, indicating habituation to this numerosity in the child, and then slides with 3 dots were shown. In another control condition, the stimuli showed to the child were more numerous (4 vs 6). The results revealed a longer fixation time by the infants for the more numerous stimuli in the first condition, but not in the second one. The authors therefore concluded that 6-7-months-old infants would be able to discriminate and represent an exact number of items, provided that this is less than 4, and that no dishabituation can occur in the control condition, because children were unable to perceive the numerosity of a set that contains more than three elements. A few years later, Antell and Keating (1983) re-proposed this study with infants from 1 to 12 days old showing vary similar results.

Of course, the ability of children and infants to discriminate numbers is not limited to the visual presentation of stimuli, because they can access the abstract representation of the number regardless of the way it is presented. In further research (Starkey, Spelke, & Gelman, 1990), 6-, 7-, and 8-month-old infants observed a projection positioned to their right, which showed two common objects and one to their left in which there were three objects. Simultaneously with these visual stimuli they heard two or three auditory signals (rolls of drums). It was observed that the newborns looked longer at the slide where the number of objects was equal to that of the perceived sounds, demonstrating that they were able to grasp the number of sounds and compare it with the number

of objects in front of them. Therefore, already in the first year of life, children understand the numerical correspondence between sets of elements presented in different sensory modalities. This implies that they possess an abstract and amodal concept of number, that is, numerical representation is not linked to the mode of presentation of stimuli.

Critically, Xu and Spelke (2000) using habituation (Figure 1.3), showed that the ability to discriminate between different quantities is present from early months of life even for numerosities larger than 4. In particular, they showed that 6-month-old infants discriminate between large sets of objects on the basis of their numerosity, when non-numerical variables are controlled, as long as the sets to be discriminated differ by a large ratio (i.e., 1:2, or 8 vs. 16 dots).

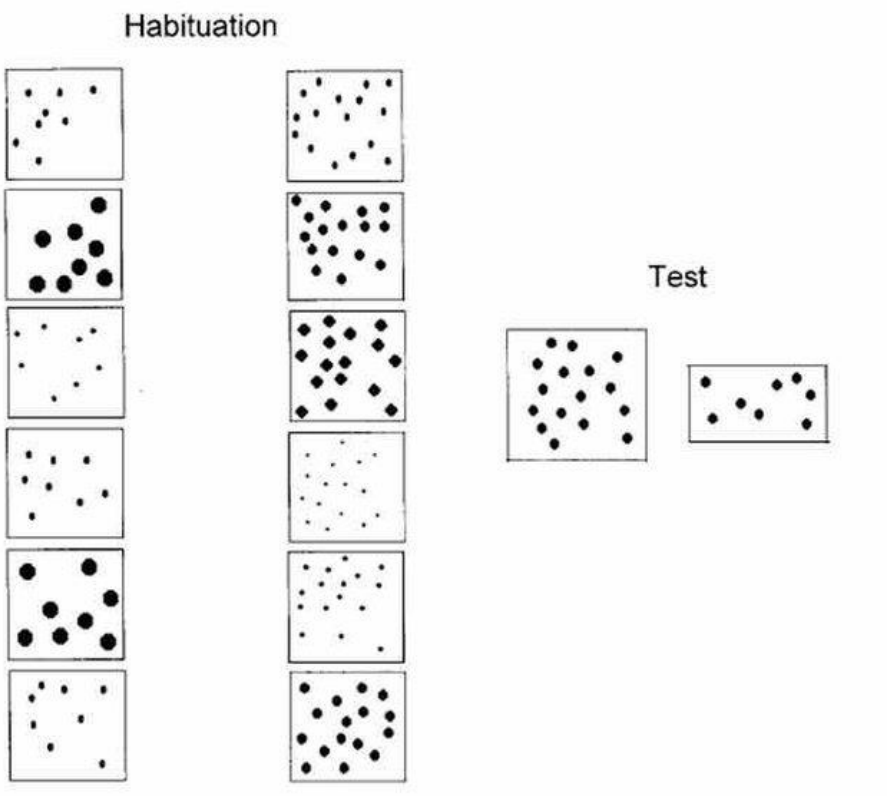


Figure 1.3: Example of stimuli used in the experiment by Xu and Spelke (2000).

All these studies suggest that infants and children in the preverbal age already possess an ability to represent the numerosity, but also a highly inaccurate representation, requiring a

minimum ratio between 2:3 and 1:2 (e.g., they successfully distinguish 8 vs 16 elements, but not 8 vs 12). However, developmental studies show that by age of 9/10 months, there is an improvement in accuracy and infants are able at discriminating even between 8 and 12 elements (Lipton & Spelke, 2003). Finally, the results obtained from all these developmental studies suggest the existence of two different numerosity representation systems: the Object-Tracking System (or Subitizing in the adult literature), which operates with numbers below 3 and 4 and the Approximate Number System, which represents large numbers.

1.3. Psychophysical and Neurocognitive signatures of the Number Sense

1.3.1. Cognitive Models of Numerosity perception

The way in which numerosity-selective neurons are tuned to a specific numerosity and how they are activated within cortical processing is explained by two theoretical cognitive models proposed to describe the extraction of quantitative information from sensory inputs.

One of the first proposed, is the "control-mode" model (Figure 1.4, left panel) hypothesized by Meck and Church (1983). This model basically suggests that each item is encoded by a pulse of a pacemaker, which is added to an accumulator. The quantity in the accumulator, at the end of the count, is then feed into memory and it constitutes an abstract numerical representation of a set of items. Thus, it is assumed that quantity is encoded by summation coding, i.e., the monotonically increasing and decreasing response functions of the neurons (see also the network model by Zorzi & Butterworth, 1999).

Another model is the "numerosity-detector" model proposed by Dehaene and Changeux (1993), which assumes the existence of circuits that normalize the visual or acoustic input with respect to their size (Figure 1.4, right panel). This model suggests the existence of three different modules: a "retina" that captures inputs of different sizes and in different positions, an "object location map" of the objects and a map of "number detectors". First, each (visual) stimulus is coded as a local Gaussian distribution of activation by topographically organized input clusters (simulating the retina). Next, items of different sizes are normalized to a size-independent code. At that stage, item size, which is initially coded by the number of active neurons on the retinotopic map (quantity code) is now encoded by the position of active clusters on a location map (position code). Clusters

in the location map project to every unit of downstream summation clusters, whose thresholds increase with increasing number and pool the total activity of the location map. The summation clusters finally project to numerosity clusters. Numerosity clusters are characterized by central excitation and lateral inhibition so that each numerosity cluster responds only to a selected range of values of the total normalized activity, i.e., their preferred numerosity. Because the numerosity of a stimulus is encoded by peaked tuning functions with a preferred numerosity (causing maximum discharge) this mechanism is termed a labeled-line code. A similar architecture was proposed by another computational model of Verguts & Fias (2004) using a backpropagation network.

The two models differ in important aspects: the model from Meck and Church (1983) operates in a serial manner and assumes representations of cardinality on a linear scale (e.g., the numerical representations are created with a process similar to counting) whereas the model of Dehaene and Changeux (1993), instead, encodes the numerosities in parallel and represent them on a logarithmic scale. However, both models assume an analogical representation of the number and predict that numerical discrimination should follow Weber's law. Finally, both models require summation units that accumulate number in a graded fashion prior to feeding into numerosity detectors at the output.

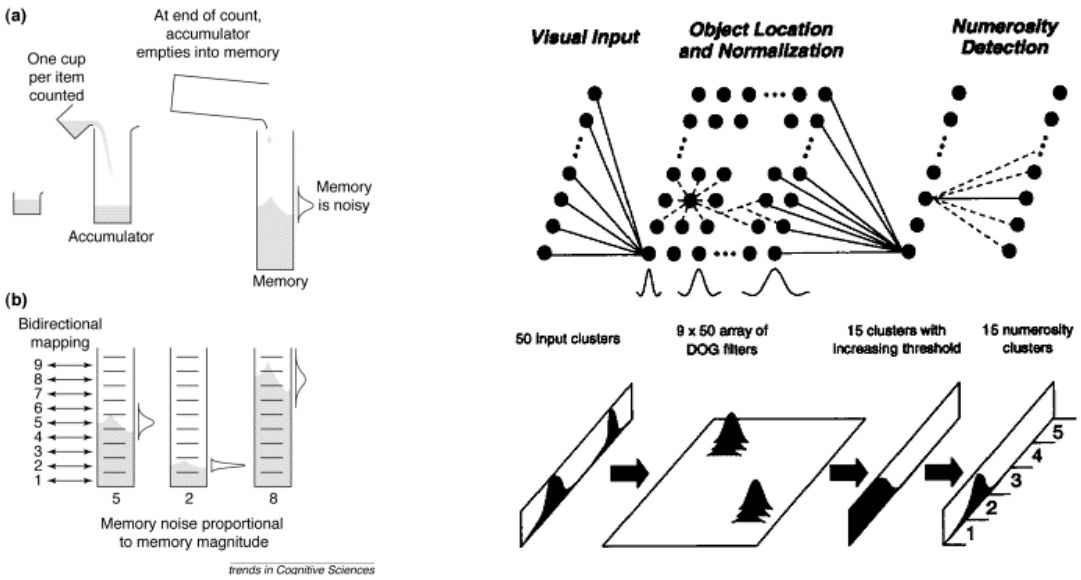


Figure 1.4: Models of Meck and Church (left) and Dehaene and Changeux (right) for the extraction of numerical information: the first foresees that the cognitive system accumulates signals obtained from each stimulus and that at the sum of the inputs is assigned a numerical label (Gallistel & Gelman, 2000). The second shows how objects of different sizes are normalized in a size-independent code and how activations are summed to provide an estimate of the number of inputs (Dehaene & Changeux, 1993).

1.3.2. Psychophysical evidence

As we briefly reported above, while numerosity representation and its precision within the range of subitizing (3-4 items) is extremely accurate, the representation of larger quantities is approximate, and the accuracy of the representation depends on the numerical size of the collection. In particular, the perception of large numbers (e.g., larger than 4 items) follows two very specific psychophysical mechanisms. The first one is the *distance effect*, whereby the reaction times and errors in comparison between numerical quantities decrease when the numerical distance between them increases. Indeed, when two numbers are distant, for example 6 and 9, observers make fewer errors and are faster in selecting the larger number than when the two numbers are very close, for example 8 and 9. The *size effect*, on the other hand, is the ability whereby, for a constant numerical distance, the difficulty in discriminating two numbers increases as their size increases. Therefore, for a constant distance, the discrimination performance of the numerosity worsens as the value of the numbers increases. For example, it is easier to judge the larger number between 6 and 7 than between 8 and 9 (e.g., Moyer & Landauer, 1967).

These properties indicate that ANS obeys the well know psychophysical Weber's law, whereby the discrimination between two magnitudes depends on their ratio. For example, the accuracy in discriminating 12 vs 24 iems (ratio 1:2) would be similar for 48 vs 96 items (ratio 1:2). On the basis of several behavioral and neuroimaging studies it has been shown that the dependence of the performance on the ratio between numbers is the main signature of ANS (e.g., Cantlon, Brannon, Carter, & Pelphrey, 2006).

The distance effect and the size effect suggest that the internal non-verbal numerical representations of the number could be organized along a mental number line (see also *Paragraph 1.6*). Two mathematical general formulations of this hypothetical line have been proposed, (Dehaene & Changeux, 1993): the *linear* model with *scalar* variability and the *logarithmic* model with *fixed* variability (see Figure 15).

The first model represents the mental number line as a series of equidistant Gaussian distributions with increasing expansion or noise. According to this view, the number is represented linearly in the mind and the variability of a given numeric representation increases linearly (e.g., scalar variability) as the value of the represented number increases (e.g., Gallistel & Gelman, 2000). Hence, in this model, the numerical representation would be linear and the distance effect and the size effect, according to this hypothesis, would be due to the increase in the "noise" of the coding

signals associated with the numerical representation. Higher values of magnitude would therefore result in greater coding "noise".

An alternative explanation is the logarithmic model, which represents numerosities as partially overlapping Gaussian curves on a logarithmic scale characterized by a fixed amount of noise (variability). According to this view, the ANS encodes the numerosities as analog quantities that can be modeled as equal Gaussian distributions with increasing overlapping as the magnitude of the numerosity increases. This theory was supported by Dehaene and Changeux (1993), who proposed a simple neural network for detecting numerosity which would presuppose a logarithmic encoding of the number, and, more recently, by Nieder and Miller (2003) who directly recorded the behavioral and neural response curves of two monkeys performing a numerical discrimination task between two sets of objects. According to this model, hence, the mental representation of the numerosity is logarithmically compressed (Dehaene, 2003). That is, the representations of a group of 12 items and of a group of 13 would be represented on a hypothetical mental line as closer than those of a group of 2 and 3 items. According to this model, the closer two representations of numerosity are represented, the more difficult it becomes to discriminate them.

Both models, however, predict similar behavioral findings since difficulty in discrimination between two numerosities, is regulated by the amount of overlapping between the Gaussian distributions (e.g., Dietrich, Huber, & Nuerk, 2015). This should explain why numerosity discrimination depends on the ratio between numerosity (e.g., the amount of overlapping between two numerosities should be similar for equal ratios).

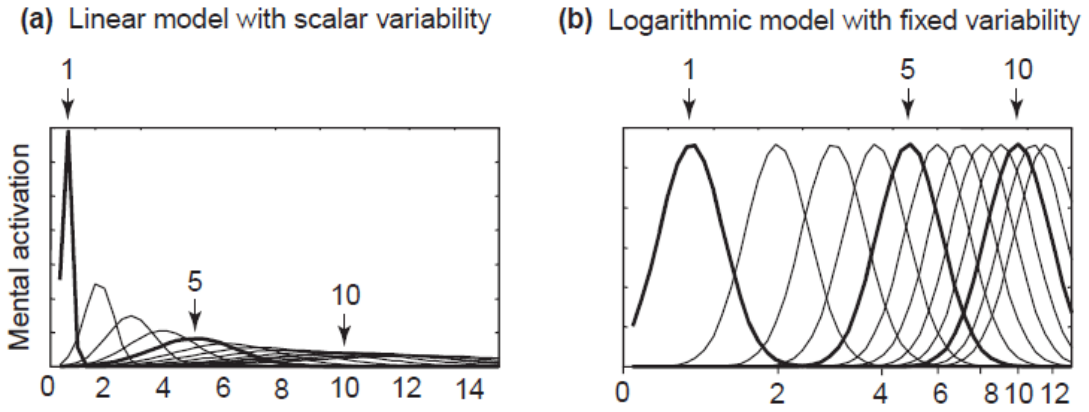


Figure 1.5: Models of the mental number line: (a) the linear model and (b) the logarithmic model.

Another important psychophysical evidence of numerosity encoding comes from studies that used the adaptation-technic. When a particular physical feature undergoes sensory adaptation (i.e.,

a kind of aftereffect), it is generally inferred the existence of a dedicated neurocognitive system encoding that specific feature (Thompson & Burr, 2009). For instance, Burr and Ross (2008) showed that exposure to an array of objects shifted the perceived numerosity of the subsequently array presented to the region that had been adapted (Figure 1.6). Accordingly, it has been recently suggested that numerosity acts as a primary sensory attribute, since visual input related to numerosity undergoes adaptation as many other primary sensory attributes such as color, motion and spatial frequency.

Furthermore, it has been shown that numerosity adaptation generalizes across sensory modalities and formats (Arrighi, Togoli, & Burr, 2014). Indeed, in the study of Arrighi, Togoli and Burr (2014) it was shown that the perception of the number of both visual and sound sequences was largely influenced by previous exposure to numerical stimuli cross-modally presented. In particular, the authors showed that numerosity of both auditory and visual sequences was greatly affected by prior adaptation to slow or rapid sequences of events. Crucially, adaptation generalized across modalities, from auditory to visual and vice versa. In addition, adaptation also generalized across formats: adapting to sequential streams of flashes affected the perceived numerosity of spatial arrays. According to the authors, the presence of adaptation to numerical stimuli in cross-modal and cross-method mode would demonstrate the domain-specific nature of a specialized system for numerical representation.

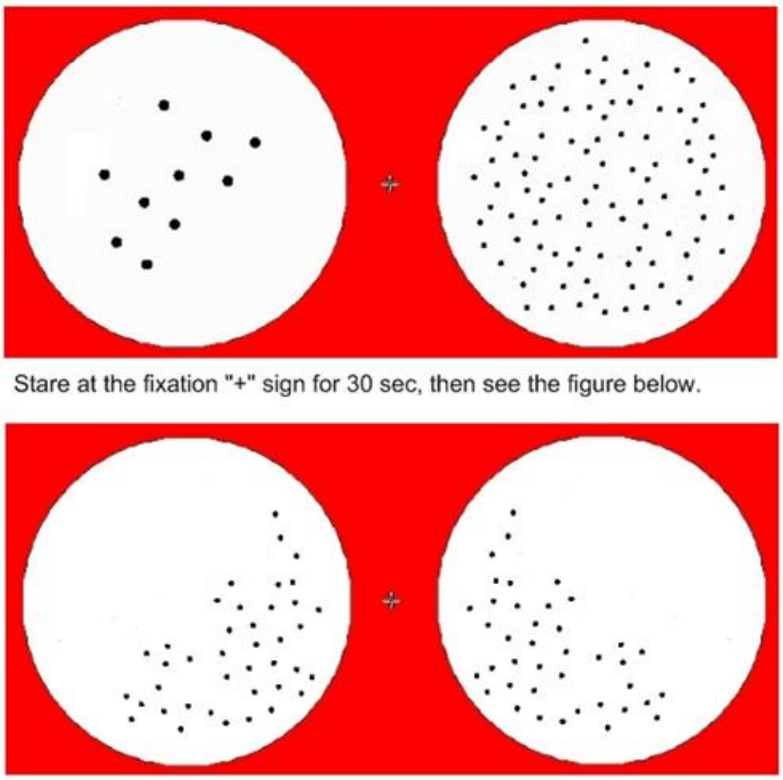


Figure 1.6: An illustration of the adaptation technic used by Burr and Ross (2008). After staring at the figure above for 30 seconds, the left side of the display should be experienced as more numerous than right, although they are actually identical.

1.3.3. Neurophysiological evidence

The first clinical evidence concerning the localization of calculation and counting processes at the neural level comes from classical neuropsychological studies of patients who had a brain injury. Acquired calculation deficits (acalculia) occurred after brain damage near the parieto-occipito-temporal junction (Henschen, 1919; Gerstmann, 1940) or in the frontal lobe (Luria, 1966). Through the observation of these cases, it has been deduced that patients with parietal lesions may have a deficit in the processing and judgment of numerosity, while other cognitive abilities remain intact (Dehaene & Cohen, 1997). In line with this observation, Deahene, Piazza, Pinel, and Cohen (2003) identified 3 neural circuits responsible for numerosity processing. This Triple Code Model assumes that lesions in one of these three circuits would determine different cognitive impairments. A first circuit in the bilateral intraparietal system would be associated with a general quantification system, the damage of which would result in a semantic deficit of the numerical domain; a second circuit localized in a region of the left angular gyrus would be associated with the verbal processing of numbers, potentially producing impairments in verbal production; a last circuit would located in the posterior superior parietal system would be involved in spatial and non-spatial attention processing, and lesions of this area would result in attention deficit in exploring the mental number line (Dehaene et al., 2003).

With the advent of non-invasive brain-imaging techniques it was possible to explore the processing of numbers in humans also in non-pathological or normal conditions. Using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), two methods for recording changes in brain metabolism that are related to neural activity, the primary role of the parietal cortex in numerical processing (whose activation is automatic and independent of numerical notation) has been confirmed (Dehaene et al., 2003). The data obtained collected over the years confirmed the existence of specific neurons tuned for numbers, which have fundamental properties shared by humans and primates and are located in comparable brain regions in the two species. At the neural level, the processing of a number therefore consists of a specific neural activation (measured with the firing rate or neuronal metabolism) of a particular population of neurons for that particular number (peak) and neighboring numbers. This activation of neighboring neurons coding specific numbers would decrease as the distance between numbers to be processed increases. Consequently, there is a greater overlap of neural representations between two neighboring numbers than between two distant numbers. Furthermore, this overlap increases as the number increases, so it is easier to compare two very distant numbers compared to two

neighboring numbers (distance effect). Similarly, it is easier to compare two relatively small numbers than two large ones, even when their numerical distance is constant (size effect).

The first fMRI studies focused on the neural correlates of mental arithmetic support the involvement of the bilateral parietal regions (Castelli, Glaser, & Butterworth, 2006). More precisely, the area of the intraparietal sulcus (IPS) is constantly activated during simple operations or numerical comparisons, performed by adult subjects (Pinel, Dehaene, Riviere, & LeBihan, 2001). Therefore, Piazza, Izard, Pinel, LeBhian and Dehaene (2004) elaborated an experimental design in which subjects were repeatedly presented with matrices of dots with fixed numerosity (e.g., 16) to habituate the population of neurons encoding this value (e.g., reduced neural activation after several exposures). After habituation, a different number from the previous ones was presented. The fMRI results revealed that only two regions, namely the horizontal segments of the intraparietal sulcus (hIPS) of the right and left hemisphere, responded to changes in *non-symbolic* numerosity and increased their activation in relation to the ratio between numerosities, regardless of the direction of the change. From this experiment emerged that the properties of these number-tuned curves in humans were very similar to those observed in the posterior parietal cortex (and frontal) cortex of primates (Nieder & Miller 2004).

The activation of the hIPS in number processing has been replicated many times, with stimuli of different nature and different paradigms (Piazza, Mechelli, Price, & Butterwoth 2006). Many studies have recorded an activation of this area, always bilateral but greater in the left hemisphere than in the right, only in response to numerical stimuli (Pinel, Dehaene, Riviere, & LeBihan, 2001). Furthermore, it has been observed that the intraparietal activation is greater during operations with large numbers (Kiefer & Dehaene, 1997), it is modulated by the numerical distance in the comparison tasks between numerosity (Pinel et al., 2001) and that it is independent of the method of presentation of the numerical input (Piazza, Mechelli, Butterworth, & Price, 2002). In a study in which subjects observed characteristics such as numerosity and other physical properties of visual and auditory events, it was found that the right hIPS was active whenever subjects focused on the numerosity of elements, regardless of how stimuli were presented (Piazza et al., 2002).

Hence, hIPS would have a central role in the amodal representation and manipulation of quantity, as it would encode the *abstract* quantitative meaning of numbers

Using recent advanced fMRI recordings on humans, it was possible to identify specific fine-grained neural regions particularly responsive to the numerosity of sets of objects (Harvey et al., 2013), such as the posterior superior parietal lobule. From the study by Harvey et al. (2013), indeed

emerged that there is an ordered topographic map connected to the processing of non-symbolic quantities in this part of the brain: the medial and lateral regions show a preference for low and high numbers, respectively. This numerical selectivity can be observed in the left hemisphere and also in the right, albeit with a lower variance and a less clear and consistent topographic structure (Figure 1.7). However, in addition to the role of parietal area for the abstract representation of numerosity, the support of other brain regions in the processing of numerosity has also been hypothesized. Tudusciuc and Nieder (2007) emphasize that the frontal lobe also contributes to the processing of non-verbal quantitative representations, even in the absence of explicit task demands. Furthermore, activation in the frontal lobe seems to be particularly prevalent in children (Ansari & Dhital, 2006). Finally, recent electrophysiological and neuroimaging studies suggest that a larger network, including the early visual cortex, might be involved in *non-symbolic* numerosity processing (DeWind, Park, Woldorff, & Brannon, 2019; Fornaciai, Brannon, Woldorff, & Park, 2017; Fornaciai & Park, 2018; Park, DeWind, Woldorff, & Brannon, 2015).

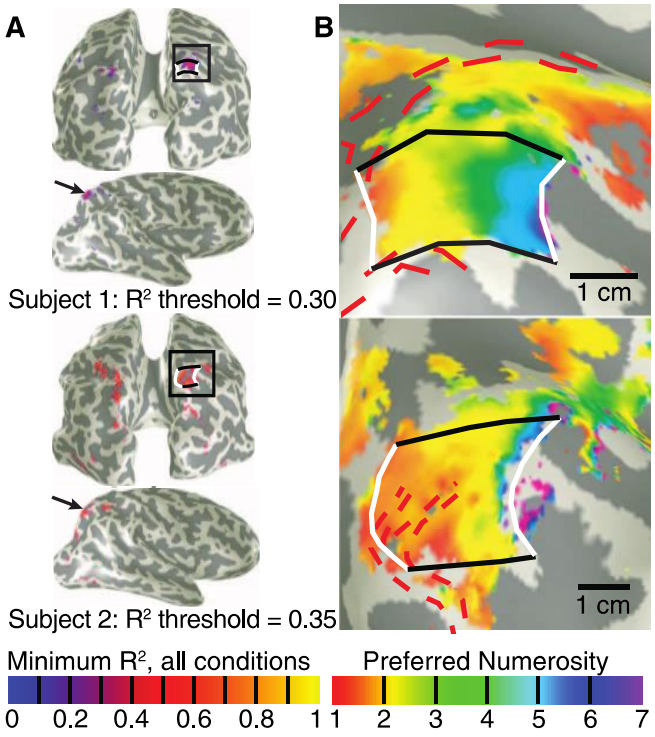


Figure 1.7: Topographical representation of the numerosity. (A) The model by Harvey et al. (2013) identified a region in the right parietal cortex in which neural populations are tuned to numerosity in all stimulus conditions (Fig. 7A). The part inside the black square is highlighted more in the enlargement in figure 7B. (B) Representation of the preferences of numerosity, based on the average of the data in all stimulus conditions, from the medial to the lateral extremities (white lines) of the region of interest.

1.4. One or two systems of Numerosity perception?

As we briefly discussed in the introduction, the existence of two distinct preverbal numerical mechanisms is not shared by all the researchers in the field. Indeed, some authors maintain that there is only one system for the estimation of numerosity; that is, that numerical perception is based on a single mechanism of representation for both small and large numerosities (e.g., Gallistel & Gelman, 2000; Meck & Church, 1983). Indeed, Cordes, Gelman, Gallistel, and Whalen (2001) asked subjects to respond to the visual presentation of a certain number in Arabic code by pressing a key as many times as the value of the number presented was. They found that mean number of presses increased as a power function of the target number, with a constant Coefficient of Variation (an index of the Weber fraction or numerical precision) within and outside the subitizing range (i.e., 1-4) suggesting that the representation of small and large numbers was based on a single mechanism.

Studies conducted on primates, in which the ability of rhesus monkeys to apply a learned numerical rule with the values from 1 to 9 items was tested (i.e., to put the numbers in order from the smaller to the larger), also reached the same conclusion. Indeed, as was reported, monkeys generalized learned rules to large values, such as 10, 15, 20 and 30 items (Cantlon & Brannon, 2006). This performance was then compared to that of human adults, who were tested in the same task. Monkeys performed qualitatively and quantitatively similar to humans and were able to represent very large values based on Weber's law. The fact that monkeys were able to extend this rule to larger numbers led the authors to conclude that they did not process numbers below ten qualitatively differently from larger numbers. According to Cantlon and Brannon, therefore, there is only one single non-verbal numerical representation system shared by monkeys and humans that allows to represent and compare the numerosities according to Weber's law.

Even in studies conducted with neuroimaging techniques conflicting results were collected. Piazza et al. (2002) used PET to identify the neural substrates of subitizing and numerical estimation mechanisms. In this experiment, the subjects had to estimate dots of various numerosities (1-4 and 6-9). The results showed the presence of a common neural network for the two mechanisms, which included the medial areas of the occipital lobe and the intraparietal areas. These data therefore seem to confirm the view of the existence of a single preverbal number system.

However, the later study by Revkin and colleagues (Revkin et al., 2008) investigated behaviorally the presence or absence of two distinct systems dedicated to the representation of numbers in the subitizing range and outside this range. These authors hypothesized that if the

subitizing system was based on a numerical process similar to that envisaged for ANS, the performance in numerical naming tasks for groups of 1 to 8 elements should have been very similar for discrimination of groups of 10 to 80 elements. The results showed a clear violation of Weber's law since the authors found a much higher precision over numerosities 1–4 than over numerosities 10–40. According to these authors, the results would bring evidence of a system dedicated to the processing of quantities within the subitizing range, hence supporting the hypothesis of two distinct numerical processing systems.

In addition, an ERP study conducted on human adults (Hyde & Spelke, 2009), which involved the use of non-symbolic numerical arrays in a condition of passive observation, recorded the activation of different brain areas for processing of small and large numbers. Specifically, an early-evoked component (N1), observed over widespread posterior scalp locations, was modulated by absolute number with small, but not large, number arrays. In contrast, a later component (P2p), observed over the same scalp locations, was modulated by the ratio difference between arrays for large, but not small number.

Therefore, although currently most authors agree with the existence of two preverbal mechanisms for processing the numerosity (i.e., ANS and subitizing), a definitive separation between the two systems has not been reached. An alternative intriguing hypothesis could be that subitizing is activated in some contexts for the processing of small numbers, but in other experimental contexts the ANS could also take charge of the processing of numerosity in the subitizing range (e.g., Agrillo, Piffer, Bisazza, & Butterworth, 2015).

1.5. Number Sense and mathematical achievements

Recent evidence has hypothesized that the ability to approximate numerical sets and to perceive numerosity is a perceptual skill comparable to the ability to perceive the color of an element (i.e., how much an element is "red" instead of "blue"), or its physical size (e.g., Piazza et al., 2010). In particular, in the same way that some people have greater difficulties in the perception of certain chromatic tones, some people would have greater difficulty in distinguishing two numerical sets with a ratio that is instead discriminable for other subjects. Hence, according to these authors, these individual differences in numerical acuity could be at the basis of the deficit in the acquisition of numerical and mathematical skills, the so-called dyscalculia, a specific learning disability that is typically diagnosed during the first years of primary school.

Interestingly, an important longitudinal study by Halberda and colleagues (Halberda et al., 2008), analyzed numerical acuity (Weber Fraction) in subjects of 14 years old, comparing their results in computerized non-symbolic tasks with those collected on them in standardized mathematical tests starting from kindergarten. The goal of this study was to explore the link between the non-symbolic numerical acuity and mathematical abilities achieved over the years. To test numerical acuity, subjects were presented with visual stimuli composed of dots of two different colors (i.e., yellow and blue). To measure their ability in mathematics tests, subjects were assessed annually (starting from kindergarten) with two standardized mathematical tests (i.e., TEMA-2 and WJ-Rcalc). The results collected showed that current individual differences in non-symbolic numerical acuity correlated with the results obtained annually for both test batteries. More precisely, the numerical acuity measured at age 14 would retrospectively predict the mathematical results collected during kindergarten (Figure 1.8). These results, according to the authors, highlight the close correlation between numerical non-verbal acuity and the ability to process and handle symbolic numerical information (but see Wilkey & Ansari, 2019).

Therefore, two different theoretical interpretation have been proposed to account for the deficits in the acquisition of numerical and arithmetic concepts. The first one favor a domain-general origin of dyscalculia whereby weakness in cognitive functions supporting the numerical processing, such as working memory, visuo-spatial abilities or phonological skills would be largely responsible for the arithmetical learning disability (Geary, 1993). The second approach, on the contrary, postulates a domain-specific origin of dyscalculia, hypothesizing that it derives directly from a deficit

in the preverbal systems supported by the numerical networks (Dehaene, Piazza, Pinel, & Cohen, 2003).

Starting from these domain-specific hypothesis, and in an attempt to understand the nature of a possible relationship between ANS acuity and developmental dyscalculia, Piazza and colleagues (2010) conducted a study in which they compared non-symbolic numerical acuity of children with a diagnosis of dyscalculia with peers without evidence of a deficit in the domain of mathematics, to understand if a difference in numerical acuity could be at the basis of this developmental deficit. The children were asked to indicate which of two sets of points was the most numerous, without counting the elements. The results of the study seem to confirm on the one hand how numerical acuity in children without a diagnosis of dyscalculia seems to improve as development proceeds (evidence already proposed, for example, by Lipton and Spelke in their 2003 study) and on the other, children with a diagnosis of dyscalculia show a lower numerical acuity (larger Weber Fraction) compared to their peers.

According to the authors, this evidence would be confirmed and strengthened by neuroimaging studies that showed how brain areas typically involved in numerical processing (i.e., parietal cortex) present a significant functional or structural difference in dyscalculic subjects, compared to the control group (Piazza, Pinel, Le Bihan, & Dehaene, 2007). These results would therefore support the hypothesis of a fundamental role of ANS in the acquisition of successive symbolic numerical skills.

In a further study, Mazzocco and colleagues (Mazzocco, Feigenson, & Halberda, 2011) tested the hypothesis that a less efficient numerical acuity of the ANS could contribute, in children with a diagnosis of dyscalculia, to the difficulties encountered in these subjects in manipulating and acquiring symbolic numerical concepts. To test this hypothesis, the subjects involved in this study were selected and divided according to whether they had received a diagnosis of dyscalculia, and with poor academic results in mathematics or with a good command. According to the authors, lower numerical acuity should predict and therefore correlate with the performance of dyscalculic children, but not with the performance of children with low and high mathematical performance. The results of this study confirmed the hypothesis showing that children with a developmental dyscalculia have reduced numerical acuity unlike their peers with both high and low mathematical performances. Also in this case, these findings would be in line with a lower activation of the intraparietal sulcus in dyscalculic children, compared to their peers (Price, Holloway, Rasanen, Vesterinen, & Ansari, 2007). Recently, it has been shown that children with developmental

dyscalculia had poor sensitivity in processing large numerosities, but no impairment in the exact enumeration of sets within the subitizing range (Decarli, Paris, Tencati, Nardelli, Vescovi, Surian, & Piazza, 2020). Once again, the acuity of ANS seems to be involved in the acquisition of symbolic numerical concepts, suggesting that a deficit present from birth in the innate numerical processing system may be related to subsequent difficulties in the acquisition of a formal symbolic numerical knowledge.

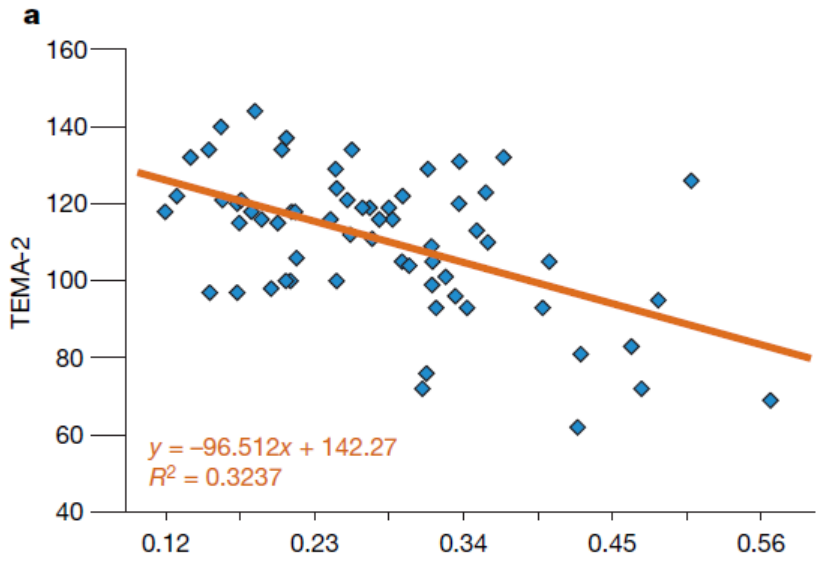


Figure 1.8: Symbolic maths achievement and the acuity of the ANS (*w*). For TEMA-2, higher numbers indicate better performance, whereas for the Weber fraction, lower numbers indicate better performance.

1.6. The link between Numerosity and Spatial processing: The Mental Number Line hypothesis

Numerical quantity processing (e.g., cardinality) and rank (e.g., ordinality) have mostly been studied separately, but the relationship between the mechanisms that underlie cardinality and serial order judgments remains indistinct (although some dissociation has been reported, e.g., Delazer & Butterworth, 1997). Hence these numerical capacities are likely two aspects of a common numerical faculty. At a behavioral level, a shared mechanism is indicated by the distance and the spatial numerical association of response codes (SNARC) effects which have been reported for cardinality and serial order processing. Indeed, in a seminal paper published almost 30 years ago, Dehaene and colleagues (Dehaene, Bossini, & Giraux, 1993) reported for the first time that adult participants tested in a parity judgment task of symbolic digits were faster to respond to small numbers with the left hand and to large numbers with the right hand. This effect has been taken as an empirical proof supporting the intuitive idea that numbers are spatially organized from left-to-right along a mental line (Galton, 1880a, b) and, beyond being replicated in various settings (e.g., Cipora, Soltanlou, Reips, & Nuerk, 2019), this phenomenon has inspired much subsequent work (de Hevia, Vallar, & Girelli, 2008). Although originally reported with symbolic stimuli (e.g., Arabic digits or number words), similar spatial-numerical associations (SNAs) have been obtained also with non-symbolic numerical stimuli (e.g., arrays of objects or sequences of tones) allowing replication in animal research (e.g., Rugani, Vallortigara, Priftis, & Regolin, 2015) and in preliterate children (Bulf, de Hevia, & Macchi Cassia, 2016; de Hevia, Girelli, Addabbo, & Macchi Cassia, 2014; de Hevia, Veggiotti, Streri, & Bonn, 2017; Ebersbach, Luwel, & Verschaffel, 2014), hence suggesting a biological foundation of this mapping. For example, the study by Rugani, Valltortigara, Priftis, and Regolin (2015, 2020) investigated the possible existence of a mental number line in three-day-old chicks, therefore virtually deprived of any prior numerical knowledge. In this study, following a training phase in which food was placed behind a panel representing a target number of elements (e.g., 5 elements), two panels containing the same number of elements were presented in the test, respectively on the right and left of the chick (Figure 1.9). The research hypothesis predicted that, if the chicks had been influenced by a hypothetical mental numerical line that orders the quantities from left to right (i.e., smaller quantities on the left and larger on the right), in the case of panels containing a number of elements smaller than the target the chicks would have preferred the left panel, and vice versa for panels with a larger numerosity. The results confirmed the hypothesis,

showing a systematic preference of the subjects towards the panels placed on the left in the "small number" conditions (ie 2 vs 2), and a systematic preference towards the panels placed on the right in the "large number" conditions (ie 8 vs 8). Overall, the results of this study indicate that a disposition to map quantities on a mental number line oriented from left to right exists independently of cultural factors and that it can be observed in animals with little non-symbolic numerical experience, supporting a phylogenetically ancient origin of this system.

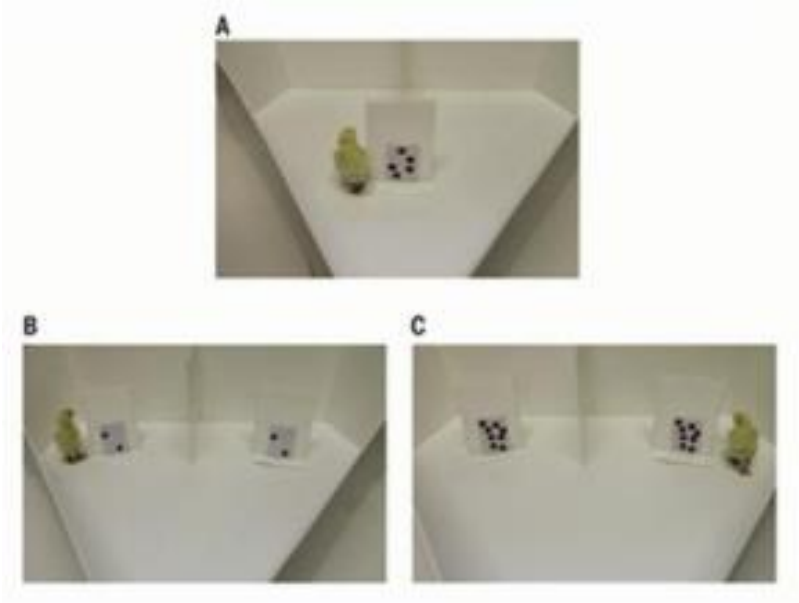


Figure 1.9: Experimental setting of condition 1 of the study by Rugani et al (2015): (a) the chicks are trained to pass a panel on which there are 5 identical elements (the target number). (b) “Small number” test phase (2 vs 2). (c) “Big number” test phase (8 vs 8).

Accordingly, recent works employing non-symbolic dot-arrays comparison tasks in adult participants reported the presence of SNAs compatible with those observed in infants and animals (Nemeh, Humberstone, Yates, & Reeve, 2018; Zhou, Shen, Li, Li, & Cui, 2016). Hence, the presence of SNAs in nonhuman animals and in humans, even at birth, has been mainly traced back to biological determinants. These include the shared representation of numerical and spatial information at the neural level (Hubbard, Piazza, Pinel, & Dehaene, 2005), as well as the early biases in the control of visuospatial attention, such as the tendency to over-attend, and start scanning from, the left side of space induced by the right hemisphere dominance for spatial processing (de Hevia, Girelli, & Macchi Cassia, 2012). The early predisposition to map number onto space would be later modeled by experiential factors, such as reading habits, which would modulate the direction of SNAs (de Hevia, 2021).

Chapter 2

2.1. The problem of Continuous Variables: The crisis of the Number Sense theory

When the number of elements to be compared or estimated is manipulated, other dimensions that are not strictly numerical generally covary with set numerosity. For example, if the physical size of the elements to be compared is kept constant between two sets, the numerically largest group will also occupy a greater cumulative surface (or is more luminant). Hence, observers can simply use this indirect cue to decide which set is numerically larger without focusing on the numerosity itself. Conversely, if the cumulative surface is kept constant between the two sets, the group with more elements will also be the one with less items. Furthermore, and contrary to mainstream model on numerosity perception which suggests that numerosity is an abstract property of the set, often the participants' performance has been found to be highly sensitive on the way in which the researchers decided to manipulate or control the continuous variables within the experimental design (e.g., Gebuis & Reynvoet, 2012a). This would therefore cast the doubt of what is really measured in numerical tasks: whether it is the ability to process numerosity or whether it is the ability to integrate different sensory information correlated or confounded with numerical information in the visual scene. This debate has led to the proposal of the so-called "indirect" theoretical models, which suggest that continuous variables play an important (if not the only) role in the development of number representation.

For instance, according to the proposal by Leibovich and Henik (2013) people would not necessarily born with the ability to represent numbers, but they would be able to develop this ability only thanks to the high correlation between numbers and continuous variables in the natural environment. In other words, this theoretical model suggests that human beings are born with the innate ability to distinguish between different values assumed by continuous variables, rather than between discrete numerosities. Over time, and thanks to the exploration of the environment they live in, children would learn to understand the high correlation between numbers and continuous variables. Only after understanding this natural correlation, children would begin to develop a more mature number sense and numerical understanding. This hypothesis takes into account the natural correlation between numerosity and continuous variables, proposing the existence of a general magnitude system capable of processing different dimensions (Leibovich et al., 2017).

Similarly, according to Gebuis and colleagues (2016), this close correlation between numerical and perceptual information would allow individuals to extract numerical information by employing only the continuous variables of the sets.

Hence, according to Gebuis and Reynvoet (2012a, b, c), assuming the existence of a specialized system for understanding numerical information would be redundant and not very economical for the brain. Since numerical and continuous variables are often strongly correlated with each other, it would appear more likely to suppose the existence of a system common to different continuous dimensions, which allows us to make numerical decisions based on continuous variables that are part of it. A holistic processing would be therefore at the core of the understanding of numerical information. In the same vein, they proposed the hypothesis that numerical information could not be extracted independently from their continuous variables, but on the contrary, during numerical processing the different perceptual "weights" associated with each continuous visual feature would be evaluated and processed to decide which group of elements contains a greater number. According to these authors, weighing the various visual cues would result in a very effective strategy in the numerical processing of non-symbolic numerical sets (Figure 2.1). Depending on the specific situation, different visual cues would be more or less informative about the number of elements (i.e., the density is more informative when it is necessary to evaluate how crowded a room is, while the convex hull turns out to be more informative when one wants to decide which of two bags contains more items).

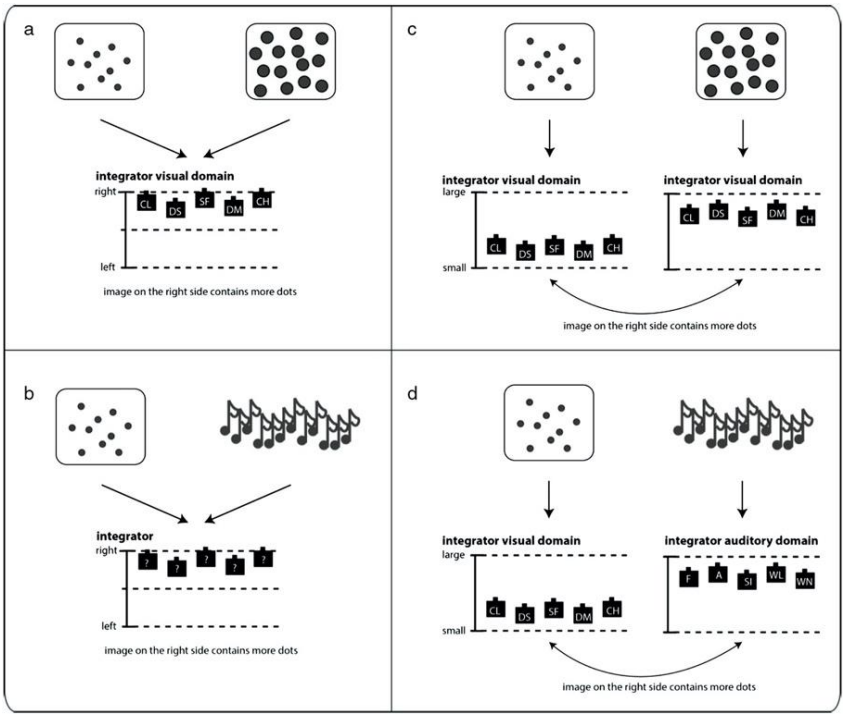


Figure 2.1: the Sensory Integration model proposed by Gebuis, Cohen Kadosh, and Gevers (2016).

Gebuis and Reynvoet (2012a) directly tested this hypothesis by using a numerical comparison task (i.e., indicate which of two sets of points was the most numerous). Participants were tested in four experimental conditions in which several continuous variables were manipulated in the arrays (i.e., convex hull, density, mean diameter of the elements, total area occupied by the elements). In some conditions, the number of elements and the values of the continuous variables were congruent, and in other conditions they were incongruent. In "congruent" trials, the panel with the greater number of elements had also higher values with respect to the continuous variable of interest (i.e., the panel with a greater number of elements also has a greater convex hull), while in incongruent trials the panel with the greater number of elements had also the smaller continuous variable of interest (e.g., convex-hull). Participants' performance showed a congruency effect induced by the visual properties of the stimuli (e.g., faster reaction times for congruent conditions compared to incongruent conditions). These congruency effects scaled with the number of visual cues manipulated, implicating that people do not extract number from a visual scene independently of its visual cues. Overall, these results seem to suggest that not only do people rely on information conveyed by the continuous variables present on the visual scene, but that such information comes from multiple sources.

In a former study by Clearfield and Mix (1999), the authors investigated in children (6 to 8 months of age) whether the early ability to discriminate different numerical sets is due to the perception of the continuous variables in the visual scene or, on the contrary, to the perception of the discrete number of elements. In particular, this study focused on the role of the total perimeter of the elements of the visual scene, as continuous variable of interest. This choice was justified by the authors on the ground that in previous studies with children, they were particularly sensitive to this type of continuous variable in the stimulus (e.g., Karmel, 1969). The total length of the perimeter of the elements was manipulated.

The children were subjected to a phase of visual habituation (Figure 2.2) in which visual panels of 2 or 3 elements (i.e., black rectangles) were presented. During the test phase, the presence or absence of dishabituation was tested with respect to two types of panels: a panel with the same number of elements presented in the habituation phase, but with a greater total perimeter; a panel with a greater number of elements, but with an equal length/total perimeter as in the habituation. The research hypothesis predicted that, if children based their numerical processing on continuous variables, dishabituation would have occurred in the condition with familiar number but with different perimeter length. Accordingly, the results suggested that the children showed dishabituation behaviors (i.e., increase in the percentage of fixations) towards the test condition familiar number but not toward the unfamiliar total length of perimeter. These results would suggest that children, in the processing of numerical information, refer to a continuous variable such as the total length of the perimeter of the elements.

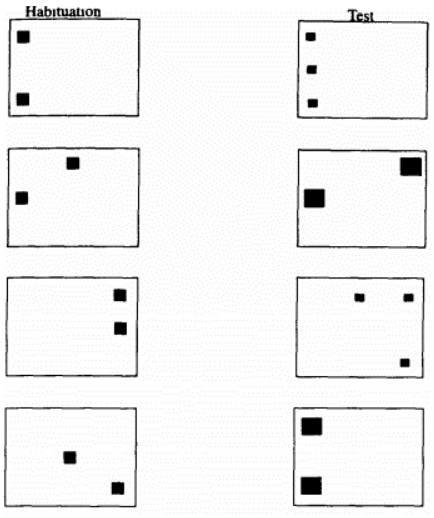


Figure 2.2: Summary diagram of the habituation phase and the test phase presented in the Clearfield and Mix experiment (1999): the two experimental conditions are manipulated through the physical size of the elements present in the test phases.

Oher models have proposed a key role played by other fundamental low-level features confounded with numerosity, mainly processed in early visual cortex (e.g., V1). For example, Durgin (2008) stressed to the role of texture density statistics (e.g., kurtosis) whereas Dakin et al. (2011) suggested that numerosity can be indirectly extracted by estimating the density of a numerical set from the visual distribution of high spatial frequencies (which captures the amount of luminance-edges) and low spatial frequencies (which captures the stimulus area) in the raw input image. Another recent visuo-biological inspired model holds that "contrast energy" can represent a surrogate feature that capture the functioning of the ANS, as adding more objects to a collection is equivalent to adding more contours or luminance-edges in the image (Morgan, Raphael, Tibber, & Dakin, 2014). Similarly, a very recent alternative theoretical proposal, named as the *brain's asymmetric frequency tuning* (BAFT) hypothesis, suggested that also SNAs would simply reflect laterality differences in the way the brain processes specific physical features in the actual numerical stimuli. In particular, this hypothesis assumes that SNAs would emerge as the result of brain asymmetries relative to the processing of the raw spatial frequencies (SF) content naturally correlated with dot numerosity stimuli (or any other visual image), with the left or right brain hemisphere preferentially dedicated to process high or low SF bands, respectively. Since small numerical arrays of dots would contain more energy in the low SF spectrum, while large arrays would have more energy in the high SF spectrum, SNAs would be the result of the lateralized SF processing in each hemisphere (Felisatti, Laubrock, Shaki, & Fischer, 2020a, b). For example, in the case of adult participants, a visually presented small array of dots would activate more the right hemisphere triggering a left-bias, hence speeding up the manual response with the left hand (Figure 2.3).

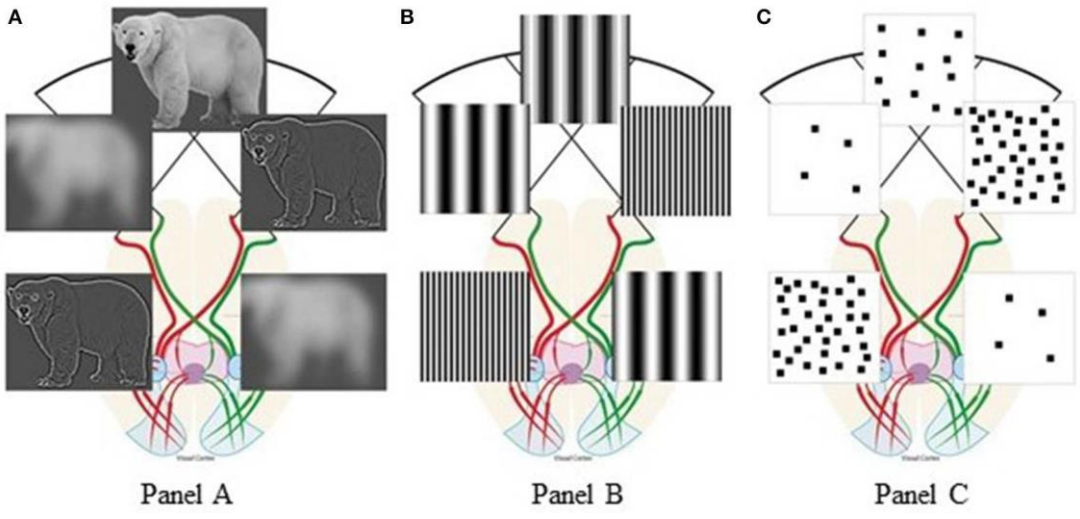


Figure 2.3: the *Brain's Asymmetric Frequency Tuning* (BAFT) proposed by Felisatti et al. (2020a, b).

2.2. Perceptual organization and Visual illusions in Numerical perception

Besides the still open controversy on the mechanisms subserving the perception of numerosity (direct vs. indirect perception), several perceptual effects linked to spatial and structural contextual factors in the visual scene suggest a certain "flexibility" of numerosity perception. Visual illusions (including size illusion such as Ponzo illusion), together with well-known grouping rules, have been used several times to investigate the perception of numerosity from the earliest years of Gestalt psychology school (for a review see Luccio, 2016). A classic study that showed how non-symbolic numerical processing can be influenced by the spatial arrangement of the elements in the scene is represented by the study of Frith and Frith, which called this effect "The Solitaire Illusion" (Frith & Frith, 1972).

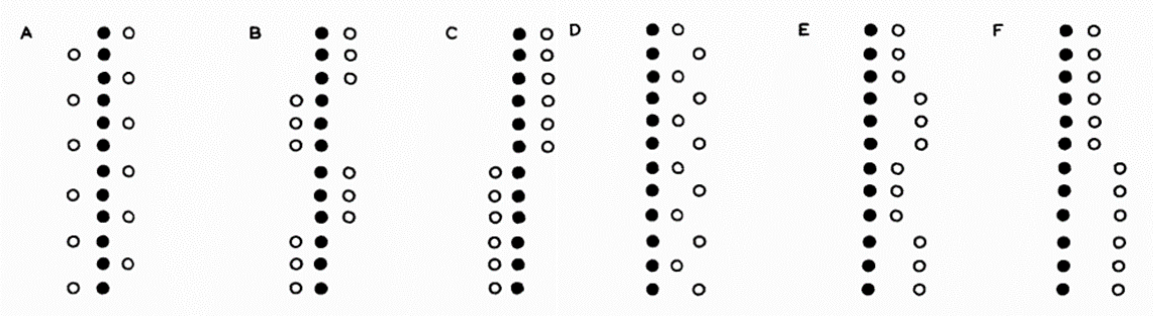


Figure 2.4: Summary scheme of the stimuli presented during Frith and Frith study (1972). In conditions A, B and C, the group of white dots is divided equally into both sides of the group of black dots, while in conditions D, E and F, the whole group is on the right side of the row of black dots (with groups that from 1, 3 to 6 elements).

In this study it has been demonstrated that a group of elements arranged in a compact and unitary whole, appears more numerous than elements placed in smaller and scattered groups. The visual scenes presented varied according to the distribution of the group of white dots (Figure 2.4). In all 6 test conditions, the group of black dots (i.e., 12 dots) were presented grouped in a vertical row, while the group of white dots (i.e., 12 dots) were placed either all on the same side of the row of black dots, or half to one side and the other half to the other side (i.e., with groups of white dots ranging between 1, 3 or 6 elements). The results show that the group of white dots was systematically underestimated in the conditions in which the two halves of elements were separated from the group of black dots (much weaker underestimation in the conditions in which

the whole group was on the same side). These results suggest that a numerical underestimation effect may derive from the spatial arrangement of the elements and, in particular, that the more scattered and less compact elements would consequently appear less numerous.

In the study by Ginsburg (1978), a possible effect of numerical underestimation with respect to sets of elements arranged in a random or regular manner was investigated. The stimuli presented were formed of elements arranged in regular distribution or in random distributions. All groups sharing the same number of elements were arranged to cover the same area in both conditions (i.e., regular group and random group). The results show a significant overestimation of the groups arranged regularly on the screen, compared to the groups arranged randomly (with the same number of elements and occupied area), demonstrating how the spatial distribution of the elements can influence numerical estimation.

Subsequently, Allik, Tuulmets and Vos (1991) proposed that the perception of numerosity would be comparable to the perception of other characteristics of the perceptual world such as the brightness. An example of the dissociation between the actual number of a set and its perception is represented by cases in which two sets of elements of equal numerosity are compared, but with a different spatial distribution. Indeed, several studies have shown that the group occupying a greater area is generally perceived as more numerous (e.g., Ponzo, 1928; Bevan, Maier, & Helson, 1963; Krueger, 1972). Perhaps the most important factor that seem to modulate the numerical underestimation and overestimation can be identified in the physical proximity between the elements of the set (Allik et al., 1991). The average proximity between elements belonging to the same set seems to be a characteristic closely associated with the perception of their number (Ginsburg & Goldstein, 1987): the greater the distance between the elements, the more numerous the group appears. An explanation for this phenomenon could be that an observer is able to process both information (i.e., numerical and spatial information), but for some reason confuses one information with the other. More likely, for reasons of cognitive economy, it becomes easier to base the numerical estimate of a set based on its most immediately processable perceptual characteristics. In an attempt to identify this perceptual characteristic used as a surrogate for purely numerical information, Vos, Van Oeffelen, Tibosch, and Allik (1988) proposed that the total area occupied by the elements influences the perception of their number. This model, called *occupancy model* (Figure 2.5), predicts that during the numerical estimate of a set, the total area occupied by its elements plays a key role. If we compare two sets of points and ask which group is the largest, the group that occupies a larger area will be chosen as the largest group (Vos et al., 1988). The idea

behind this model is that two points placed very close have a lower impact on the perception of number than two points placed at a greater distance (in a manner very similar to item proximity; see also Allik & Raidvee, 2021).

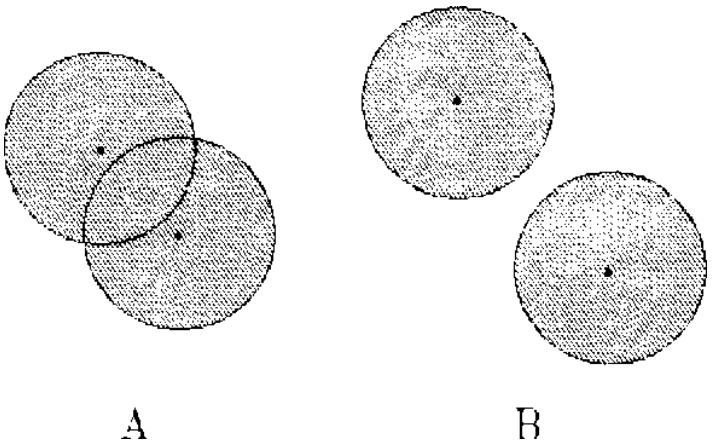


Figure 2.5: Basic principle of the occupancy model: (a) if two points are close for a distance less than their radius R, their areas overlap. (b) If the distance between the points is greater than the length of their radius R, there is no intersection between the two occupied territories. The authors hypothesized that the perception of numerosity is proportional to the totality of the area occupied by the points (Allik & Tuulmets, 1991).

In addition to the spatial effects on the perception of numerosity, there are also structural factors that influence the perception of numerical size. One of these factors is the connection between the elements. For example, in a study by Koesling, Carbone, Pomplun, Sichelschmidt and Ritter (2004) it was shown that when elements were connected to form the shape of a closed polygon, they appeared much less numerous than when they were unconnected.

Following a similar rationale, He, Zhang, Zhou and Chen (2009) investigated how physical thin lines that connect points could influence the perception of set numerosity. Participants performed a numerical discrimination task in which they compared a reference panel of 12 black points randomly distributed with test panels ranging from 9 to 15 elements. The lines present in the test panels were randomly distributed by spatial position and orientation, making possible three different levels of connection: zero connections, one connection and two connections (Figure 2.6). In general, the results show that connecting the elements into pairs using physical lines affects numerical perception. More specifically, connecting the elements into pairs, keeping as much as possible constant continuous cues, led to an underestimation of the ensemble directly proportional

to the number of connected pairs. Similarly, in the study of Franconeri and colleagues (Franconeri, Bemis, & Alvarez, 2009) it was demonstrated that connecting the elements of a scene using thin lines, reduces the perception of numerousness (underestimation). These studies, in particular, have been taken as evidence that numerosity is based on segmented objects, rather than on continuous visual features. According to the grouping principle of connectedness (Palmer & Rock, 1994), the lines should force the visual system to process the individual dots as a whole "dumbbell" object, which represent the input unit of numerosity processing.

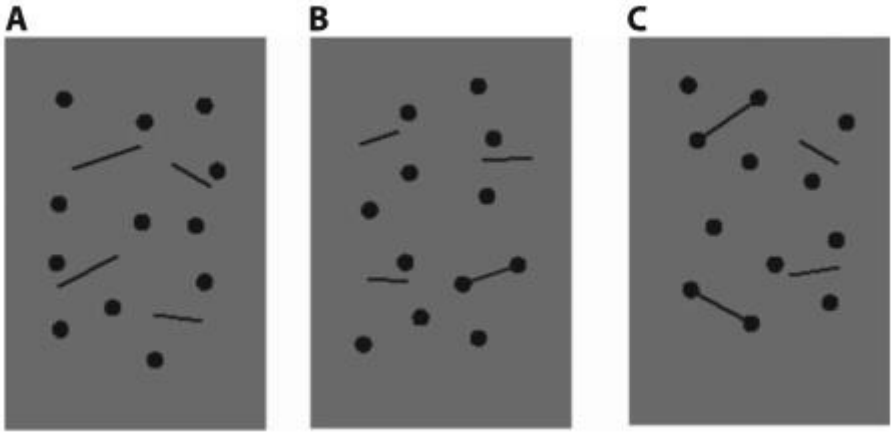


Figure 2.6: Summary illustration of the three test conditions present in Experiment 1 by He et al. (2009). (a) zero connections, (b) two connected points, and (c) four connected points.

Furthermore, some recent evidence indicates that other grouping principles may affect the numerosity processing. For instance, also Im and colleagues (Im, Zhong, & Halberda 2016) have shown that grouping the items by proximity reduces numerosity perception. More recently, Poom and collegues (Poom, Lindskog, Winman, & Van den Berg, 2019) reported that under prolonged viewing conditions, numerosity was underestimated when the items were grouped by color or motion. Similarly, homogeneously colored, shaped or oriented arrays seems overestimated compared to heterogeneous arrays (Redden & Hoch, 2009; DeWind, Bonner, & Brannon, 2020).

2.3. The present work: Visual Illusions and Fourier analysis as a tool to study Numerical Perception

Because the visual mechanisms of ANS are still largely debated, in this work we directly investigated the problem of continuous variables in numerical perception, and focused on the following question: is non-verbal numerosity processing based on segmented objects or is it rather based on a mere processing of low-level visual confounds?

To do so we took advantage of two different, but complementary, experimental approaches: the first one is based on the use of ***Visual Illusions*** to manipulate selective visual features processing, such as item segmentation or grouping strength, without altering the physical features of the stimuli. The second method is based on a computational approach exploiting ***Fourier analysis*** (and SF equalization) to process raw physical content of visual stimuli in order to make (low and high) spatial frequencies uninformative about numerosity in the scene.

In particular, in **Chapter 3** we presented a psychophysical paradigm in which connectedness effect was employed to manipulate item segmentation. Physical lines were replaced by Kanizsa-like illusory contours (ICs) lines to manipulate the grouping strength of the items in the set, controlling all the continuous features across levels of connectedness. In **Chapter 4**, we employed well known psychophysical properties of *reverse-contrast* Kanizsa-like to suppress/manipulate the illusory brightness confounds accompanying the Ics-lines (Figure 2.7). In **Chapter 5**, we investigated whether perceived numerosity and perceived continuous cues are two independent visual information affecting the task. The Ponzo illusion was used to manipulate the perceived convex-hull/density of the set while we concurrently manipulated the grouping by means of Ics lines.

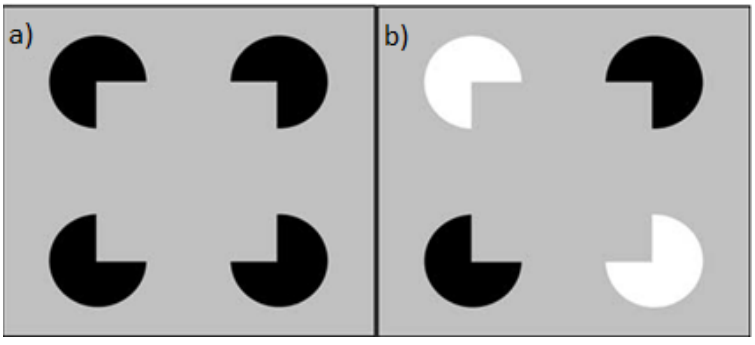


Figure 2.7: (a) classic Kanizsa square illusion. (b) reverse-contrast Kanizsa square illusion.

In **Chapter 6** we combined both a computational and a classic psychophysical approach to further show that numerosity perception is based on segmentation mechanisms. In particular, we applied Fourier analysis to equalize spatial frequency (SF) in all the stimuli (Figure 2.8) while we concurrently vary the grouping strength among items with Ics. In **Chapter 7** we also investigated whether the main signature of the Weber like encoding of numerosity (e.g., ratio-dependent performance) is not primarily explained by stimulus energy or SFs correlated/confounded with numerosity. Finally, in **Chapter 8**, we tested whether the non-symbolic spatial numerical association effect (a behavioral signature of the mental-number line hypothesis) is independent from physical SF content of the stimuli and from brain lateralization related to SF processing.

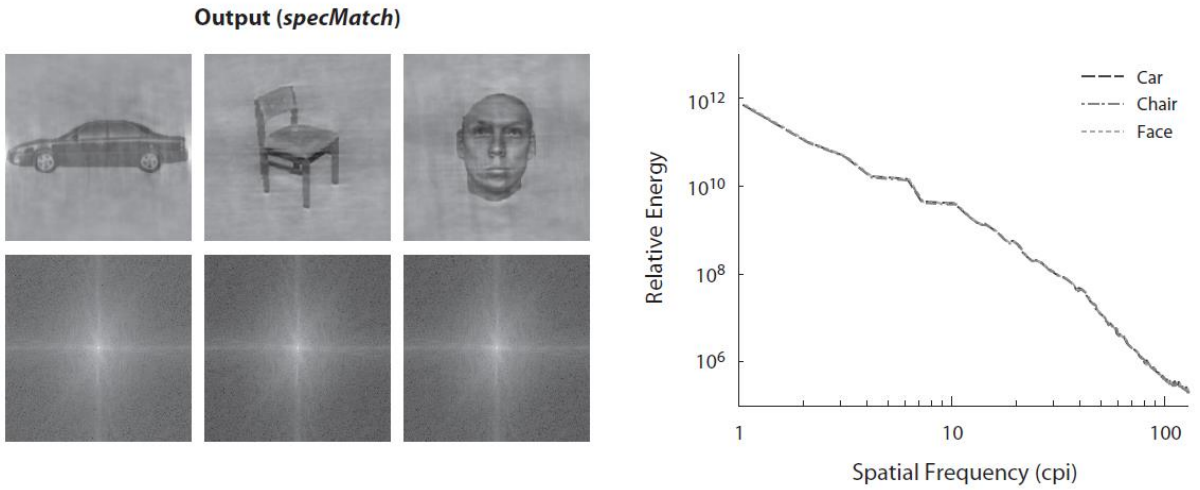


Figure 2.8: An example of the output after processing a set of generic stimuli to equalize their power spectrum.

Chapter 3

3.1. Visual illusions as a tool to hijack numerical perception: Disentangling nonsymbolic number from its continuous visual properties

Chapter adapted from: Adriano, A., Rinaldi, L., & Girelli, L. (2021). *Journal of Experimental Psychology: Human Perception and Performance, 47*(3), 423–441. https://doi.org/10.1037/xhp0000844

3.2. Introduction

As we introduced in the previous chapter (Chapter 2), visual illusions have been recently used to shed light on the debate of what are the building blocks of numerosity perception (e.g., Franconeri et al., 2009). Here, we employed an original behavioral method based on illusory contours (ICs) to disentangle whether visual numerosity processing operates over discrete units or rather over continuous variables. ICs are visual experiences of objects whose edges are not defined by luminance discontinuities with the background. The Kanizsa Triangle and the Ehrenstein Illusion probably represent the most popular examples of these visual phenomena (Grossberg & Mingolla, 1987; Wagemans et al., 2012).

In particular, in our study, we manipulated the level of connectedness of a numerical set by connecting some of the items with Kanizsa-like illusory lines: in this way, we could manipulate the degree of object-segmentation processes in the visual scene, but at the same time keeping constant all the low-level features of the image. Some previous studies, indeed, have already suggested that numerosity would not be extracted from raw low-level features (e.g., density, brightness, contours, etc.), but rather from the quantity of discrete entities of a segmented visual input. For instance, it has been shown that arrays with pairs of dots connected by task-irrelevant lines were *underestimated* compared to unconnected arrays, revealing that perceived numerosity monotonically decreased with the number of item pairs (Fornaciai & Park, 2018; Franconeri et al., 2009; He, Zhang, Zhou, & Chen, 2009; He, Zhou, Zhou, He, & Chen, 2015). The hypothesis is that, as suggested by the grouping principle of element connectedness (Palmer & Rock, 1994), the task-irrelevant lines would force the visual system to process two connected dots as a single unified

perceptual object, which would represent the input-units of visual numerical computation. This should explain the origin of the numerosity underestimation effect driven by element connectedness.

It is important to note that, in these previous studies manipulating the "connectedness" of the set, the physical connections may have involuntary introduced further methodological confounds. For instance, some authors have recently suggested that the irrelevant lines may have affected the salience of connected items tagged for enumeration, and thus the reported underestimation could simply be due to the reduced visibility of the individual items in the pair (Kirjakovski & Matsumoto, 2016). To avoid these specific confounds, Kirjakovski and Matsumoto (2016) replaced the real lines with illusory contours. Critically, ICs are perceived by both humans and animals (for a review see Nieder, 2002). They are also represented as real objects in secondary visual area (V2) of monkey and human visual cortex (Peterhans & von der Heydt, 1989, 1991) and are detected in parallel by the visual system (Davis & Driver, 1994), being therefore the ideal stimulus to replace real lines. Accordingly, a recent study using illusory contour lines showed a numerosity underestimation proportional to the number of pairs in the set, thus in accordance with studies employing real lines (Kirjakovski & Matsumoto, 2016).

However, a crucial limitation of this recent study (Kirjakovski & Matsumoto, 2016) is that low-level confounds across connectedness conditions were not fully controlled, thus leaving open the possibility that the reported underestimation could be explained by differences in items clustering and/or spatial profiles across sets (e.g., density, convex hull or occupancy index). Their stimuli generation method indeed produced stimuli that appear as more locally "clusterized" as the number of ICs increased, since in the patterns with no illusory contours (0 ICs) all the items were free to span in the space whereas in the other two conditions (2 ICs or 4 ICs) a subset of inducers was spatially constrained to be always neighbors (in order to generate the ICs). Furthermore, the size of the convex hull (e.g., the area covered by the smallest virtual polygon enclosing the items) also decreased as the number of ICs increased (see Simulation section in the *Supplementary Materials*). Consequently, the underestimation found in previous studies might be explained without the need of any visual segmentation in numerosity processing.

In the current study, to avoid all the previous potential confounds, we replaced the real lines with ICs lines to connect the items in the set, but we simultaneously controlled the spatial arrangement of the items by keeping constant their position across the connectedness conditions. In this way, we could fully rule out the role of continuous variables in numerosity perception. We,

therefore, aimed at clarifying whether visual-segmentation mechanisms affect numerosity processing in a typical comparison task and in an estimation task, which requires more precise numerosity processes than comparison judgments (e.g., Gebuis & Reynvoet, 2012b).

On these grounds, we carried out 4 experiments: in Experiment 1, 2A & 2B and 3, participants were tested in a comparison task, whereas in Experiment 4 they were tested in an explicit estimation task. In the comparison task, participants were asked to indicate the numerically larger set between two rapidly presented visual stimuli, each one presented on the left and right hemispace. One of the two stimuli was a reference set, containing a fixed numerosity (12 items), while the other stimulus was a test set with the number of items varying between 9 and 15. In Experiment 1, test sets contained pairs of collinear open inducers items that prompted 0 ICs, 2 ICs or 4 ICs lines connecting two items, thereby theoretically inducing an incremental numerical underestimation (i.e., maximum underestimation with 4 ICs). In striking contrast, test sets in Experiment 2 (2A & 2B) included 0, 2 or 4 collinear closed inducers that did not trigger analogous underestimation illusions as in Experiment 1. Crucially, in both experiments, spatial profiles of test sets were matched by duplicating each item position across the conditions, while only a subset of the items were suitably rotated to create the required experimental conditions (see the Methods section for further details about stimuli generation and the experimental design). Thus, test stimuli were fully comparable in terms of continuous variables: they all had the same items distribution (convex hull extent/shape and density), as well as cumulative surface and brightness, across the experimental conditions (e.g., Gebuis & Reynvoet, 2012a; Hurewitz et al., 2006; Katzin et al., 2020). In the Experiment 3, to investigate the possible interplay between visual segmentation mechanisms and continuous features, we concurrently manipulated both the number of ICs (0, 2 or 4) and the size of the convex hull of test patterns (e.g., the area covered by the smallest virtual polygon enclosing the items), while participants performed a comparison task preserving a similar design of the Experiment 1 and 2 (e.g., continuous features were completely neutralized across ICs levels, within each level of convex hull size). Following a similar rationale, in Experiment 4, we presented as stimuli only the test sets of the Experiment 1, which varied between 9 and 15 items and contained 0 ICs, 2 ICs or 4 ICs lines. In this specific case, however, stimuli were individually presented at the center of the screen for a brief interval time and subjects were asked to explicitly estimate the number of objects.

In sum, our experimental design and stimuli should therefore guarantee *a priori* that any findings cannot be attributed to different items distributions and spatial profiles nor by other low-level features when ICs are manipulated. Indeed, if numerosity is merely computed on the raw

amount of low-level features of an unsegmented scene (e.g., density, convex hull, etc.), then the variation of the ICs connecting-lines should not affect the numerosity judgments, since the total amount of continuous variables (or their weights) is constant across conditions. That is, the subject's performance measured by the Points of Subjective Equality (PSE, the number of items required in test sets to be judged subjectively equal to the reference set) in Experiment 1 and the mean subjective estimation in Experiment 4 should not change when the number of illusory connections was manipulated. On the contrary, if numerosity is based on single perceptual objects of a segmented visual input, then connected arrays should be underestimated proportionally to the number of ICs connecting lines (or item-pairs), although *item distributions* and *spatial profiles* are carefully matched across the connectedness levels. Thus, following this prediction, increasing the number of illusory connections should increase the PSE in Experiment 1 (e.g., more items are required in test sets to be perceived as equal to the reference, in order to compensate the reduction in perceived numerosity triggered by the illusory lines); similarly, in Experiment 4, an increment in illusory connections should decrease the mean subjective estimation. This may lead to challenge the hypothesis that the underestimation effect reported in previous studies (e.g., Kirjakovski & Matsumoto, 2016) that used ICs lines was just a by-product of continuous variables (like the convex hull or item-density), supporting the idea that the underestimation effect is actually due to the mandatory binding of the single items into unified discrete objects, which represent the input-units of numerosity processing.

Furthermore, Experiment 2 (comparison of closed inducers) was designed as a further control to test whether any underestimation effect was specifically triggered by ICs grouping or rather was due to a general effect caused by inducer edge alignments. The Experiment 2 was composed of two sub-experiments. In the first one (Experiment 2A), the inducers were closed with a thicker line (e.g., 4 pixels) as in the original paper by Kirjakovski and Matsumoto (2016), whereas in the second one (Experiment 2B) we used a thinner line (e.g., 1 pixel) in order to equalize the notch size of the inducers and their visibility in the periphery, thus resembling the visual conditions of Experiment 1.

In both experiments, we predicted that if numerical underestimation was specifically due to ICs, no underestimation effect should be found with either 0, 2 or 4 aligned closed inducers, as these features prevent a strong ICs formation (e.g., equal PSEs across conditions). On the contrary, if the effect was due to inducers alignment, we should find an increase of the PSEs as we increase the number of aligned closed inducers.

Finally, in Experiment 3, we predicted that if numerosity is based on segmented items, numerosity should be underestimated (e.g., increase in the PSE) as the number of ICs-connections increased in test patterns, although all continuous features were neutralized across connectedness levels. However, as suggested in previous work, we also predicted that numerosity should be independently underestimated (e.g., increase in the PSE) when the size of the convex hull of test stimuli was relatively smaller compared to when it was relatively larger than the reference (e.g., Gebuis & Reynvoet, 2012a, 2012b).

3.3. Experiment 1: Comparison Task with Open Inducers

3.3.1. Materials and methods

3.3.1.1. Participants

We performed an a priori power analysis with G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to determine our needed sample size. Because our study was inspired by the work of Kirjakovski and Matsumoto (2016), who found an effect of ICs with a very small sample (N=6), we decided to calculate the sample size for a moderate to high effect size. The calculation established that, in order to obtain a medium to high effect size (η^2_p = .1) with a 80% of power for the main effect of the number of ICs (0, 2 or 4 ICs) in a one-way repeated measures ANOVA (3 levels of measurements; alpha = .05), a minimum sample of 16 participants was required.

A sample of 17 participants (13 females) took part in the study. The mean age was 25.64 years (*SD* = 6.19), they were all undergraduate or postgraduate students from Milano-Bicocca University. Handedness was assessed by asking participants which hand they typically used for writing: a total of 16 subjects were classified as right-handed. All participants had normal or correct-to-normal vision. All the participants were naïve about the purpose of the experiment. Each subject signed an informed consent document before the experiment began and the study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Local Ethical Committee (protocol N° RM-2020-230).

3.3.1.2. Stimuli

The stimuli were projected on an Acer AL1716 17" LCD monitor (display area: 338 x 270 mm, 1280 x 1024 pixels, 60 Hz) connected to a pc desktop Olidata (AMD Athlon X2 240, 2.80 GHz, 4 GB ram, Windows® 7). The stimuli were randomly generated off-line by a custom Python/Psychopy script and projected by means of a Psychopy routine (Peirce, 2007). Stimuli were constructed with the same specifications as in Kirjakovski and Matsumoto (2016), but with the changes to accommodate our control of continuous variables.

The whole experimental set was composed of 168 test stimuli (56 random spatial patterns cloned across the 3 levels of connection) and 168 reference stimuli (56 random spatial patterns repeated within the 3 levels of connection) generated off-line.

The reference patterns always contained the same numerosity (*N*=12). In particular, they were composed by 12 black "pac-man" like items (diameter = 20 pixels; notch width = 4 pixels; notch length = 10 pixels, measured from the center; RGB = -1, -1, -1) spatially scattered and randomly rotated at an angle varying across 360° to avoid collinearities and the pop-out of ICs. The test patterns contained a variable numerosity (from 9 to 15 "pac-man" like items). We generated a first set of 56 test stimuli for the 0 ICs condition (8 random visual patterns were generated for each of the 7 numerosity values in test stimuli) and we coupled them to the 56 reference patterns. In each stimulus of the original 0 ICs set, the distance between the "pac-man" items that could prompt the required number of ICs for the other two connectedness conditions was randomly chosen among four possible values (center-to-center distance = 22, 25, 28, and 31 pixels). To keep constant spatial profiles of test sets across the levels of connectedness (i.e., and thus to control continuous variables), each different test pattern for each numerosity in the 0 ICs condition was cloned among the different levels of connection for 2 ICs and 4 ICs test stimuli. Thus, we kept constant the spatial position of all the single items in a given pattern from the 0 ICs set, but a sub-set of "pac-man" items was appropriately rotated and aligned to prompt the required number of ICs for the other test conditions. In this way, the 56 different reference patterns were associated with the same spatial pattern of test stimuli across the levels of connectedness.

In sum, a given spatial pattern in test stimuli was cloned across the level of connectedness for each numerosity and was associated with the same reference spatial pattern (Figure 3.1). This is done to ensure that the difference in the PSEs across ICs conditions was not due to the fact that test patterns were assigned to a different reference spatial pattern across levels (e.g., since the task was a relative magnitude judgment). Reference stimuli were constructed with the same spatial constraints for the items as the original test patterns in the 0 ICs condition.

All the single items were drawn on a grey background (RGB = 0, 0, 0) and reference and test stimuli were projected within two squared panels (240 x 240 pixels) against a black background (RGB = -1, -1, -1). Furthermore, we constrained the single "pac-man" items in test and reference stimuli to be distant at least 20 pixels from the 4 square edges and to not overlap with each other (minimum center-to-center distance = 22 pixels).

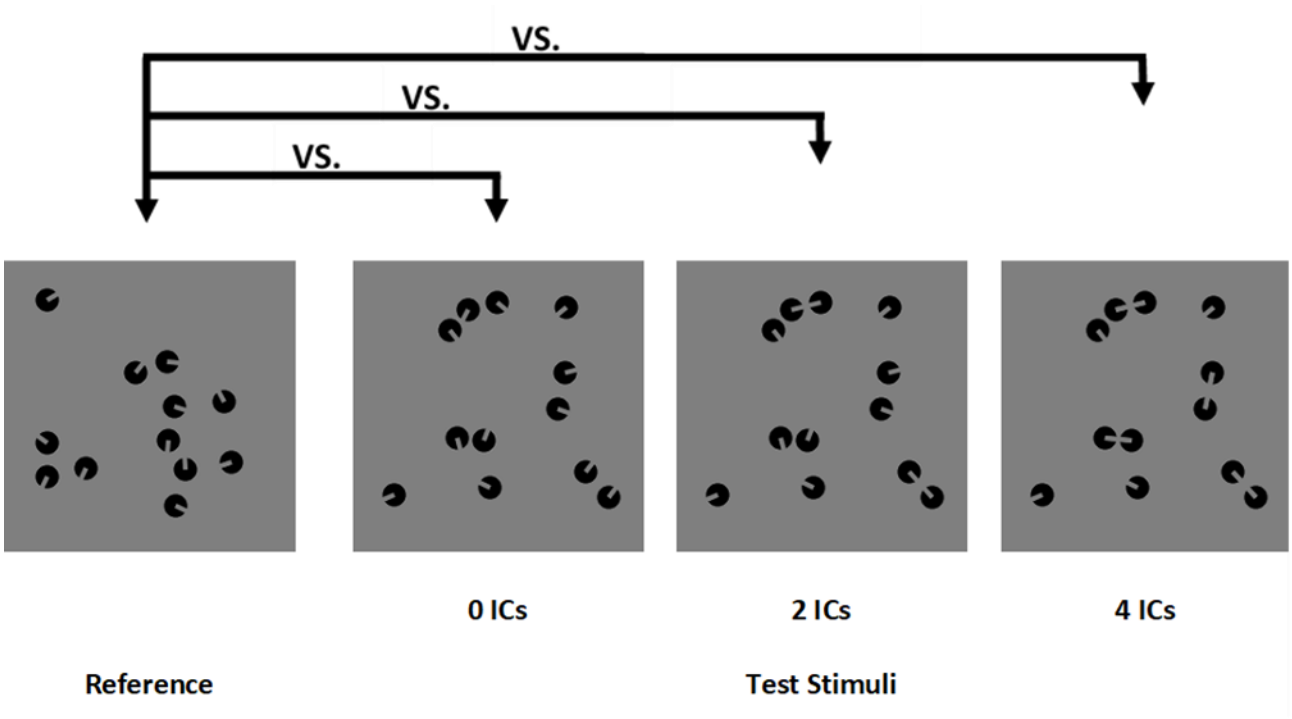


Figure 3.1: The Reference set was always composed of 12 items. Test patterns varied from 9 to 15 items and contained 0 ICs, 2 ICs or 4 ICs. All the test stimuli had the same convex hull, density and total surface when averaged across the levels of connectedness (as in the examples depicted). Black arrows indicate that a given reference pattern was presented with the same spatial pattern across the 3 levels of connectedness of test stimuli.

3.3.1.3. Procedure

The experiment was conducted in a quiet and dimly illuminated room. All the participants were individually assessed. Subjects were comfortably seated at a viewing distance of 80 cm from the screen. The general procedure was explained by the experimenter to each participant before starting the experiment and detailed instructions were also provided on the display. No information about the illusions or the connectedness was given to the subject. The participants performed a two-alternative forced choice task in which they were asked to choose the set containing more objects between two rapidly presented patterns by pressing the corresponding keys on the keyboard.

The experimental phase was preceded by a brief training composed of 24 trials to allow the subject to familiarize with the task. In the training phase, we presented only the condition with 12 items in reference vs. 9 items in test (8 trials for each test set with 0, 2 or 4 ICs). Each experimental trial started with a black background (RGB = -1, -1, -1) lasting 1000 ms, followed by a fixation cross (Font: Times; Size: 16 pixels; RGB = 0, 0, 0) projected for 1000 ms alone, and then, two panels

appeared at the left or right of the fixation cross (72 pixels between the nearest edge of each square panel to the fixation) for additional 400 ms (Figure 3.2). The side of the reference and test patterns was counterbalanced and randomized across trials. After the stimuli offset, an empty screen (RGB = -1, -1, -1) was presented until the participant's answer. The subjects could select the stimulus by pressing the appropriate key with their left or right index finger ("F" key for the left stimulus and "J" key for the right stimulus). Response time was not restricted, but we emphasized in the instructions to answer as fast as possible after the stimulus off-set. After the practice session, two counterbalanced blocks of 336 randomly ordered trials were presented, for a total of 672 experimental trials (32 trials for each of the 21 conditions), separated by a forced pause (3 min) at the half of the whole session and by 2 short forced pauses (1 min) at the half of each block. The whole experiment lasted around 40-45 min.

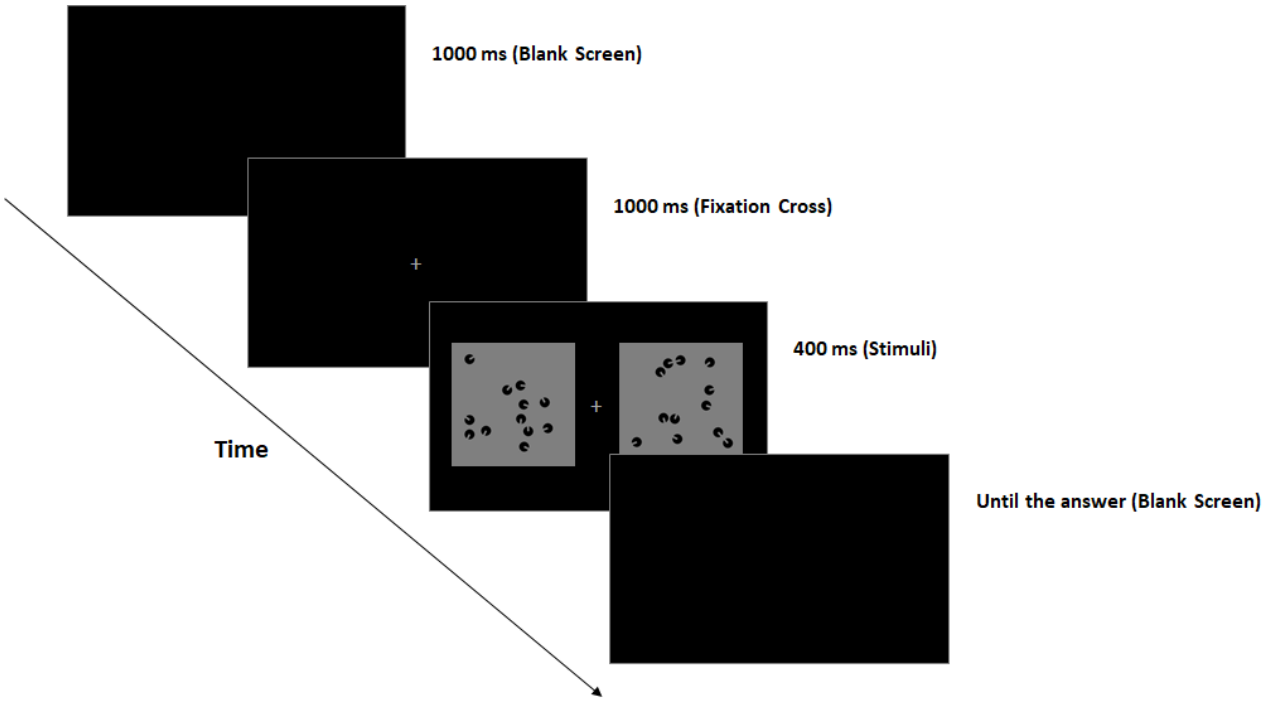


Figure 3.2: The discrimination task. Subjects had to decide which stimulus was numerically bigger by typing the relative key to specify the left or right stimulus side (F or J key). The side of reference and test pattern was balanced and randomized.

3.3.2. Results

The data were analyzed with *R-Studio* (2018, v. 3.6.2; http://www.rstudio.com/) and *Jamovi* (2019, v. 1.1.5; https://www.jamovi.org) softwares. Psychometric functions for each condition were generated by fitting Gaussian cumulative distribution functions and parameters were estimated with a parametric approach based on maximum likelihood method, using *Quickpsy* package for R (Linares & López-Moliner, 2016). Each of the 21 conditions (i.e., 3 ICs conditions x 7 numerosities) contribued with 32 trials to the psychometric function.

The dependent variable was the proportion of trials in which the test pattern was chosen as more numerous than the reference pattern. The 50% of the chosen test patterns was set as threshold level and represented the point of subjective equality (PSE): that is, the number of dots in test patterns required in order to be subjectively judjed as equal to the the reference patterns (12 items). Furthermore, as an index of the precision of the numerical discrimination and to confirm that the performance follows the Weber's law (e.g., the variability of judjments is proportional to the the magnitude of the stimuli discriminated, resulting in a constant ratio) we calculated the Coeffienct of Variation (CoV, Halberda & Odic, 2014). Following Halberda and Odic (2014), the CoV was computed as the ratio between the standard deviation (SD) and the PSE of the psichometric functions for each condition.

In order to minimize biases in estimating the psychometric function parameters, we fitted the psychometric curves taking into account the typical lapse in performance (e.g., missing a trial, finger-errors) by allowing the value of the guess rate (γ) and lapse rate (λ) parameters to vary in the default range of 0 – 0.05 (Wichmann & Hill, 2001). The 95% confidence intervals of individual PSEs were estimated running 200 bootstrap resampling of the data. Thus, for each subject we calculated the PSE (See Supplementary Materials for the individual psychometric functions and their goodness-of-fit; respectively, Figure S3.3 and Table S3.4) and the CoV of the fitted functions for the 3 levels of connectedness and the overall sample data were submitted to two separate one-way repeated measures ANOVAs with the number of ICs (0, 2, or 4) as within factor and either the PSEs or the CoV as dependent variables. When sphericity assumption was violated, we applied the Greenhouse-Geisser epsilon (ε) correction and we reported the original *F*, *df* and corrected *p*-values.

As can be visually observed in the figure plotting the psychometric functions obtained pooling over the aggregate data of the whole sample (Figure 3.3A), we found a rightward shift of the psychometric curves for 2 and 4 ICs conditions, thus suggesting an underestimation of the

perceived numerosity as we increased the number of connections in test stimuli (please note that this graph is reported to illustrate the technic, but all subsequent analysis was done with similar functions over individual subjects). The analysis of individual PSEs showed a significant effect of the number of ICs, $F(2, 32) = 7.03$, $p = .003$, $\eta^2_p = .30$. That is, the PSEs increased with the number of ICs (Figure 3.3B). Such a pattern, was confirmed statistically by means of post-hoc comparisons (Bonferroni correction), revealing a significant difference between 0 ICs and 4 ICs, $t(32) = -3.75$, $p = .002$, while no significant difference was found between 0 ICs and 2 ICs, $t(32) = -1.96$, $p = .17$, and between 2 ICs and 4 ICs, $t(32) = -1.79$, $p = .24$. Furthermore, a polynomial trend analysis shows a significant linear trend only, $t = 3.78$, $p < .001$.

Furthermore, the analysis of the CoV of the psychometric functions for the three levels of connectedness showed no significant differences across conditions, $F(2, 32) = .20$, $\varepsilon = .65$, $p = .72$, $\eta^2_p = .013$, suggesting an equal numerical estimation precision across ICs conditions as predicted by the Weber's law (See Supplementary Materials; Figure S3.1). Individual CoVs were also correlated to probe that individual differences in number processing precision remained stable across ICs conditions: results showed a positive, strong relationship (See Supplementary Materials for more details; see also Figure S3.2 and Table S3.3).

In addition to frequentist analyses we also ran Bayesian ANOVAs over both the PSEs and the CoVs with the number of ICs as independent variable. These additional analyses confirmed the main results reported here (See Supplementary Materials for more details; see also Table S3.1 and Table S3.2).

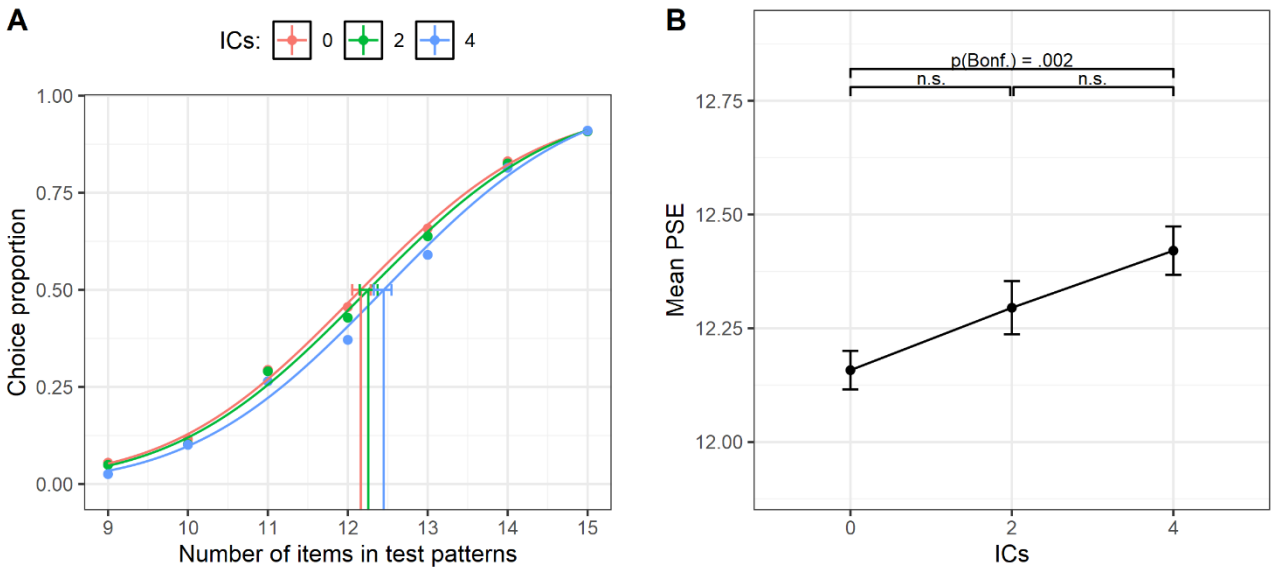


Figure 3.3: (A) Psychometric functions obtained fitting Gaussian cumulative distribution function (for 0-2-4 ICs) pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs and y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

3.3.3. Discussion of Experiment 1

The results of Experiment 1 suggest that patterns with 2 or 4 ICs required more items, compared to the condition with 0 ICs, to be judged as equal to the reference pattern when the actual number of items between the two sets was the same. Indeed, the PSEs monotonically increased with the number of ICs in the set: in other words, numerosity in test sets was underestimated when the "pac-man" items were rotated to prompt the ICs. This increase in the PSEs indicates that more items are required for compensating the decrease of perceived numerosity caused by the grouping of few individual items into pairs. Yet, such findings may be as well explained by a general effect due to inducers edge alignment rather than to the completion of the illusory lines (cf. Kirjakovski & Matsumoto, 2016). To exclude this possibility, in Experiment 2, the "pac-man" shapes were the same as in Experiment 1 but, this time, each of them was closed with a line to prevent ICs. Previous studies showed that closing the notch with a line strongly reduced the formation of Kanizsa-like ICs (Peterhans & von der Heydt, 1989, 1991; von der Heydt, Peterhans, & Baumgartner, 1984). If the results of Experiment 1 were due to inducers edge alignments, rather than to the ICs lines, we should find an equal increase in the PSEs even in Experiment 2.

The Experiment 2 was run in two different versions. In the Experiment 2A the line closing the inducers was of the same size (e.g., 4 pixels thick) as in the original paper by Kirjakovski and Matsumoto (2016), whereas in the Experiment 2B the size of the closing line was of only 1 pixel thick, in order to maintain the notch of the inducers as much as possible similar to the notch of the open inducers.

3.4. Experiment 2: Comparison Task with Closed Inducers

3.4.1. Experiment 2A: 4-pixels closing line

3.4.1.1. Materials and methods

3.4.1.1.1. Participants

Since this study replicates the design and the power analysis of the Experiment 1, a new sample of 18 undergraduate and postgraduate students (mean age ± sd = 24.72 ± 2.67 years, 15 females, 16 right-handed), with normal or correct-to-normal vision, was recruited for the second study. All the subjects were naïve regarding the purpose of the experiment.

3.4.1.1.2. Stimuli and Procedure

The design, stimuli parameters and their generation method as well as the procedure were identical to the Experiment 1. The only difference in the stimuli generation was that inducers were closed with a curved line (4 pixels thick), thus completing the overall circular shape of each item. The notch of the inducers thus became 6 pixels length and was –as in Experiment 1– 4 pixels wide (Figure 3.4). For the sake of clarity, we adopted the same labeling as in Experiment 1. Hence, to define the number of collinear pairs in test sets, we named the conditions 0 "ICs", 2 "ICs" and 4 "ICs" as in the previous experiment.

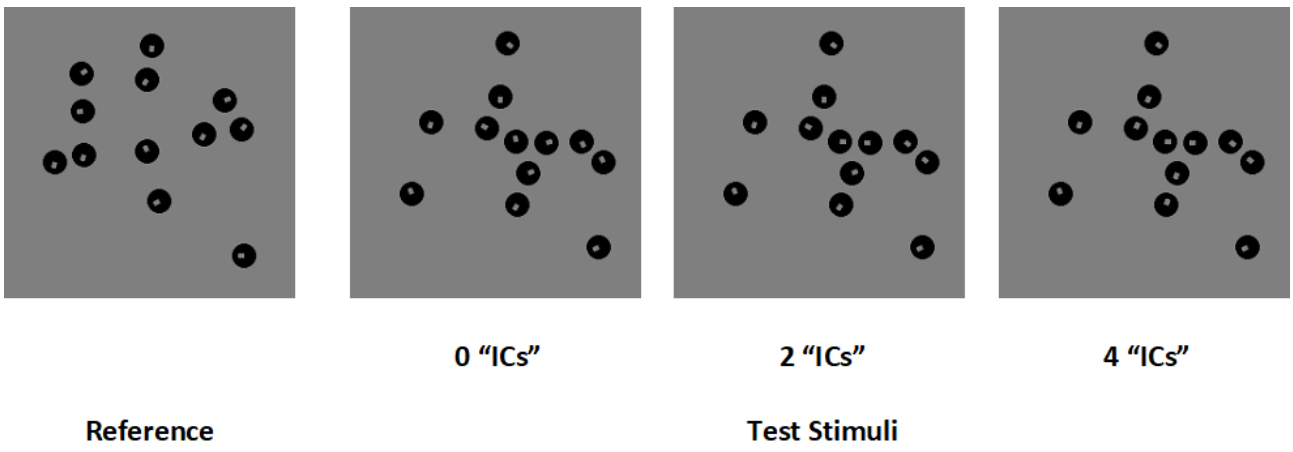


Figure 3.4: Reference was always composed of 12 items. Test patterns varied from 9 to 15 items and contained 0 "ICs", 2 "ICs" or "4 ICs". All the test stimuli had the same convex hull, density and total surface when averaged across the conditions. As in Experiment 1, a given reference pattern was presented with the same spatial pattern across the 3 conditions of test stimuli.

3.4.1.2. Results

Data were analyzed with R-Studio and Jamovi. Individual psychometric functions were generated fitting Gaussian cumulative curves following the same procedure adopted in Experiment 1 (See Supplementary Materials for the individual psychometric functions and their goodness-of-fit; respectively, Figure S3.6 and Table S3.8). Two separate one-way repeated measures ANOVAs were carried out with either the PSE or the CoV of the psychometric functions as dependent variables and the number of "ICs" (0, 2 or 4) as independent variable. The analysis showed no significant differences across the conditions, as the number of "ICs" did not affect the PSEs, $F(2, 34) = 1.28$, $p = .29$, $\eta^2_p = .07$, (Figure 3.5). No significant linear or quadratic trend was observed (all $p > .05$). Similarly, no effect of condition was found for the CoV, $F(2, 34) = 2.56$, $p = .09$, $\eta^2_p = .13$, (Figure S3.4). As in Experiment 1, individual CoVs were also correlated to control that individual differences in number processing precision remained stable across conditions: results showed a positive, strong relationship (See Supplementary Materials for more details; see also Figure S3.5 and Table S3.7). Finally, we also ran supplementary Bayesians statistics over both the PSE and the CoV (See Supplementary Materials for more details; see also Table S3.5 and Table S3.6), which overall confirmed the statistical patterns reported here.

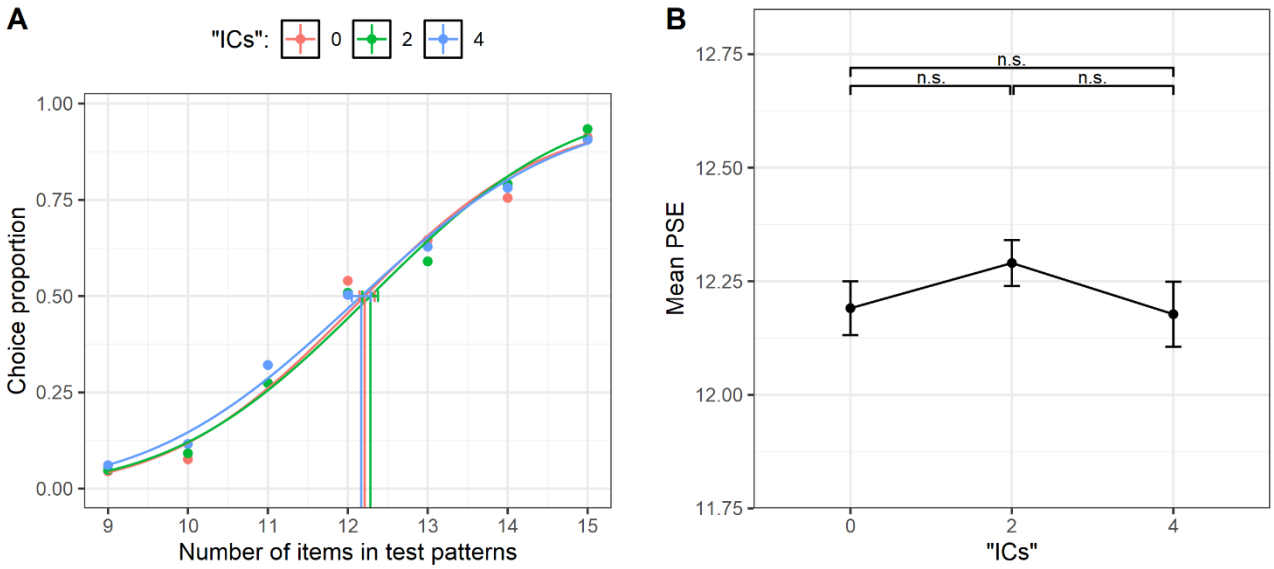


Figure 3.5: (A) Psychometric functions obtained fitting Gaussian cumulative distribution function (for 0-2-4 ICs) pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs and y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

3.4.2. Experiment 2B: 1-pixel closing line

3.4.2.1. Materials and methods

3.4.2.1.1. Participants

A new sample of 18 participants (mean age ± sd = 27.16 ± 6.38 years, 10 females, 16 right-handed), with normal or correct-to-normal vision, was recruited and performed on-line this study through Pavlovia, a repository and launch platform allowing on-line PsychoPy experiments (www.pavlovia.org). All the subjects were naïve regarding the purpose of the experiment.

3.4.2.1.2. Stimuli and Procedure

The design, stimuli parameters and their generation method as well as the procedure were identical to the Experiment 2A. The only substantial difference in the stimuli generation was that the inducers were closed with a curved line of 1 pixel thick (instead of 4 pixels; see Figure 3.6). Due to COVID-19 social restrictions in Italy, this study was performed by means of an on-line Psychopy routine.

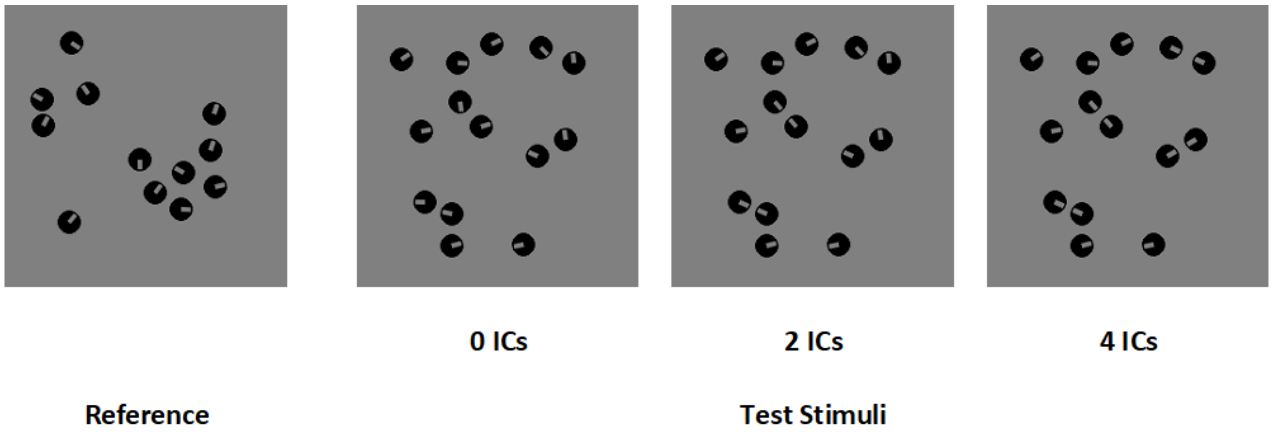


Figure 3.6: The reference stimulus was always composed of 12 items. Test patterns varied from 9 to 15 items and contained 0 "ICs", 2 "ICs" or "4 ICs". All the test stimuli had the same convex hull, density and total surface when averaged across the conditions. As in Experiment 1 and 2A, a given reference pattern was presented with the same spatial pattern across the 3 conditions of test stimuli.

3.4.2.2. Results

Data were analyzed in analogy with Experiment 1 and Experiment 2A (see Figure S3.9 and Table S3.12 for the individual psychometric functions and their goodness-of-fit, respectively). Two separate one-way repeated measures ANOVAs were carried out with either the PSE or the CoV of the psychometric functions as dependent variables and the number of "ICs" (0, 2 or 4) as independent variable. The results showed no significant effect of the number of "ICs" over the PSE, $F(2, 34) = 0.73$, $p = .48$, $\eta^2_p = .04$, Figure 3.7. No significant linear or quadratic trend was observed (all $p > .05$). The same analysis was run for the effect of the number of "ICs" over the CoV. We found no significant effect of the number of "ICs" over the PSE, $F(2, 34) = 1.39$, $p = .26$, $\eta^2_p = .07$, Figure S3.7. No significant linear or quadratic trend was observed (all $p > .05$). Bayesian analysis overall confirmed the frequentist analysis (Table S3.9 and Table S3.10). As in Experiment 1 and 2A, individual CoV were highly correlated across conditions (Figure S3.8 and Table S3.11).

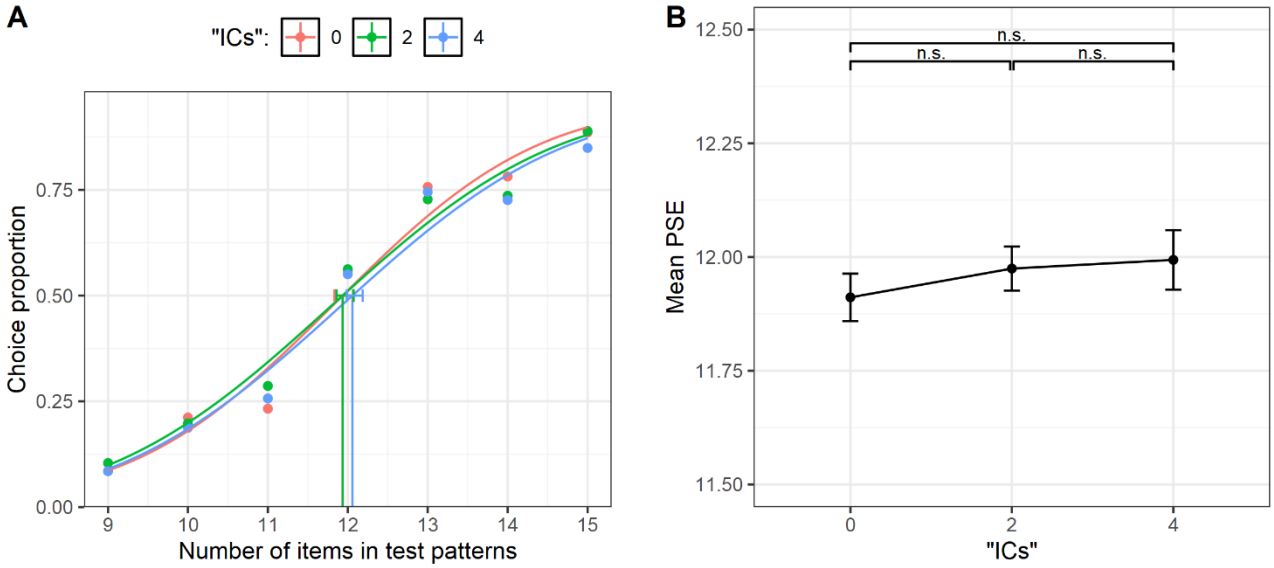


Figure 3.7: (A) Psychometric functions obtained fitting Gaussian cumulative distribution function (for 0-2-4 ICs) pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs and y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

3.4.3. Discussion of Experiment 2A and Experiment 2B

The results of both Experiment 2A and Experiment 2B indicate that the mere alignment of edges between a few inducers was not sufficient to group the items into a perceptual object and to produce a numerosity underestimation; this in contrast with the pattern observed in Experiment 1.

The same results were obtained when inducers were closed with a line of the same size (e.g., 4 pixels thick) as used in Kirjakovski and Matsumoto (2016), and when the closing line was reduced to the minimum size (e.g., 1 pixel thick) in order to equalize the size and the visibility of the notch alignment in visual periphery.

This result further corroborates the possibility that the underestimation effect found in Experiment 1 is specifically due to the binding of the single items in the pairs driven by the modal ICs lines. On these grounds, we designed a further experiment (Experiment 3) in order to investigate wether both discrete and continous information may affect the numerical perception. That is, in Experiment 3 we varied both the number of ICs and the size of the convex hull of test patterns. In this way, we investigate the possible additional or interactive role of continuous variables, which has beeen largely supported by previous research.

3.5. Experiment 3: Comparison Task with Open Inducers and small/large Convex Hull

3.5.1. Materials and methods

3.5.1.1. Participants

Due to Covid-19 restrictions in Italy, the participants were recruited through Pavlovia, a repository and launch platform allowing on-line PsychoPy experiments (www.pavlovia.org). A new sample of thirty-four subjects was recruited (21 females, 26 right-handed) with a mean age of 28.79 years (SD = 7.5). They were all naïve regarding the experiment.

3.5.1.2. Stimuli and Procedure

The stimuli were projected by means of an on-line Psychopy routine (Peirce, 2007) and all the experimental materials (stimuli, etc.) were downloaded before to start the experiment from the Pavlovia repository to a temporary local folder stored on the computer of each participant. The stimuli, the experimental design and the task adopted were similar to those of Experiment 1. The only difference was that, in order to manipulate the size of the convex hull, the surface of the virtual polygon enclosing the items was specifically constrained in all the stimuli sets. In particular, all the stimuli in the test patterns for each numerosity (9 to 15) and ICs (0, 2 or 4) condition were generated with 2 different standard levels of convex hull (i.e., constraining the area of the virtual polygon within a predefined value). Thus, half of the 8 random spatial patterns for each numerosity and ICs condition were generated with a relative larger area of the convex hull (area range: 18.1×10^3 - 18.3×10^3 px^2) and the other half with a relative smaller area (area range: 12.7×10^3 - 12.9×10^3 px^2) as compared to the reference stimuli convex hull area (area range: 15.4×10^3 - 15.6×10^3 px^2). In sum, within each block, half of the test stimuli had a relatively smaller convex hull area and the other half a relatively larger convex hull area (i.e., as compared to the reference).

As in the Experiment 1, each pattern was cloned across IC levels and the same cloned patterns were associated to the same reference pattern within each level of convex hull (Figure 3.8). The experiment was thus composed of 2 equal blocks of 336 trials (672 total trials) separated by a short break at the half of the session.

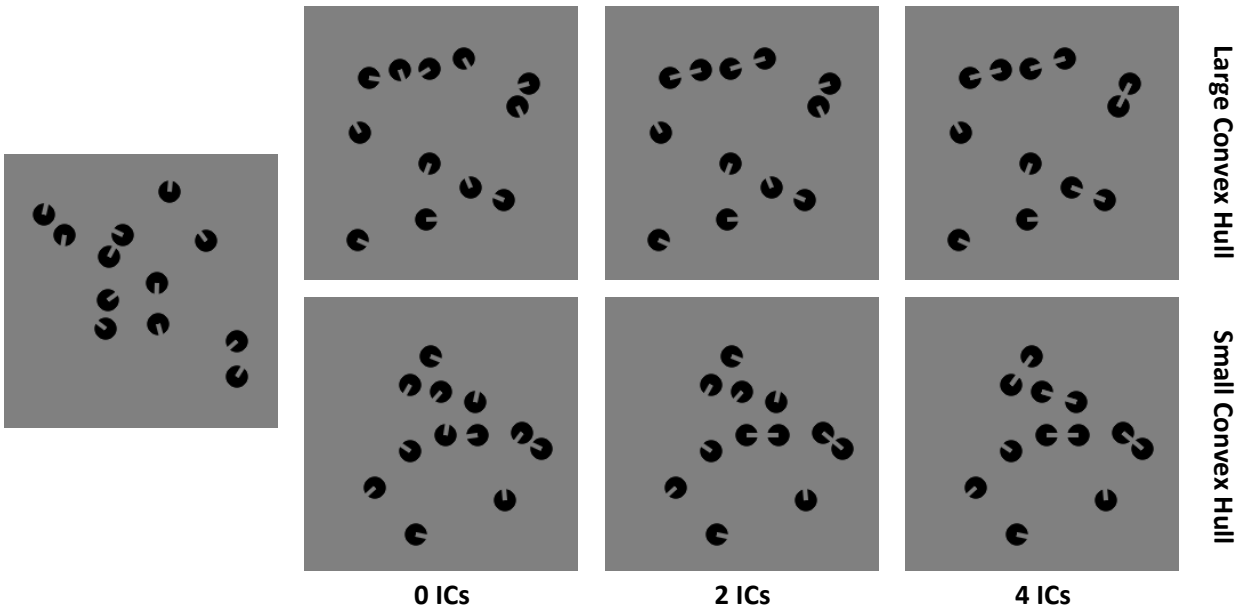


Figure 3.8: The reference stimulus was always composed of 12 items. Test patterns varied from 9 to 15 items and contained 0 ICs, 2 ICs or 4 ICs, spatially enclosed within two different convex hull size (large or small). All the test stimuli had the same convex hull, density and total surface when averaged across the same convex hull size level. As in Experiment 1, within each level of convex hull size a given reference pattern was presented with the same spatial pattern across the 3 conditions of test stimuli.

3.5.2. Results

Data were analyzed with R-Studio and Jamovi. As in the previous experiments, for each subject we calculated the PSE and the CoV of the fitted functions[1] (See Supplementary Materials for the individual psychometric functions and their goodness-of-fit; respectively, Figure S3.12 and Table S3.16) for the 6 experimental conditions and the overall sample data were submitted to two separate 3 x 2 repeated measures ANOVAs, with the number of ICs (0, 2, or 4) and the convex hull size (small, large) as within factors and either the PSEs or the CoV as dependent variables. The Greenhouse-Geisser epsilon (ε) correction for violation of sphericity was applied when needed and original *F*, *df* and corrected *p*-values were reported.

As can be observed in Figure 3.9A there was a rightward shift of the psychometric curves when ICs were manipulated, with both large and small convex hull size. The repeated measures ANOVA with the individual PSEs as dependent variable showed a significant main effect of the number of ICs, $F(2, 66) = 8.36$, $\varepsilon = .79$, $p = .002$, $\eta^2_p = .20$, Figure 3.9B. Post-hoc comparisons (Bonferroni correction), revealed a significant difference between 0 ICs and 4 ICs, $t(66) = -4.0$, $p < .001$, and between 0 ICs and 2 ICs, $t(66) = -2.73$, $p = .02$, while no difference was found between 2 ICs and 4 ICs, $t(66) = -1.27$, $p = .62$. Furthermore, a polynomial trend analysis shows a significant linear trend only, $t = 4.0$, $p < .001$.

We also found a significant main effect of the convex hull size, $F(1, 33) = 4.57$, $p = .039$, $\eta^2_p = .12$, As can be observed in Figure 3.9B, the PSEs increased when we reduced the convex hull of the test patterns, thus suggesting that numerosity is underestimated when the convex hull size decreases. No interaction between ICs and convex hull size was found, $F(2, 66) = .06$, $p = .94$, $\eta^2_p = .002$. A similar 3 x 2 repeated measures ANOVA was performed with the CoV as dependent variables. Results showed that the main effect of the number of ICs was not significant, $F(2, 66) = 2.26$, $\varepsilon = .84$, $p = .12$, $\eta^2_p = .06$. However the main effect of the convex hull size was significant, $F(1, 33) = 18.8$, $p < .001$, $\eta^2_p = .36$. Indeed, the CoV was larger in the small convex hull size condition, suggesting that numerical precision decreases when the convex hull size decreases (Figure S3.10). The interaction between ICs and convex hull size was not statistically significant, $F(2, 66) = .037$, $\varepsilon = .83$, $p = .94$, $\eta^2_p = .001$.

[1] Notice that psychometric functions of one subject (Subject n. 9) had a very poor fit in only one condition out of six (e.g., deviance $p < .01$). Data were analyzed with and without this subject, but the overall pattern of results did not change. For this very reason, we opted for including this subject in the analyses presented here.

Individual CoVs were also correlated to probe that individual differences in number processing precision remained stable across conditions. Results showed a positive strong relationship across all the conditions, in analogy with previous experiments (See Supplementary Materials for more details; see also Figure S3.11 and Table S3.15).

Finally, in addition to the frequentist analyses, we also ran Bayesian ANOVAs over both the PSEs and the CoVs with the number of ICs and the convex hull size as within factors. These additional analyses confirmed the main results of the frequentist analyses (See Supplementary Materials for more details; see also Table S3.13 and Table S3.14).

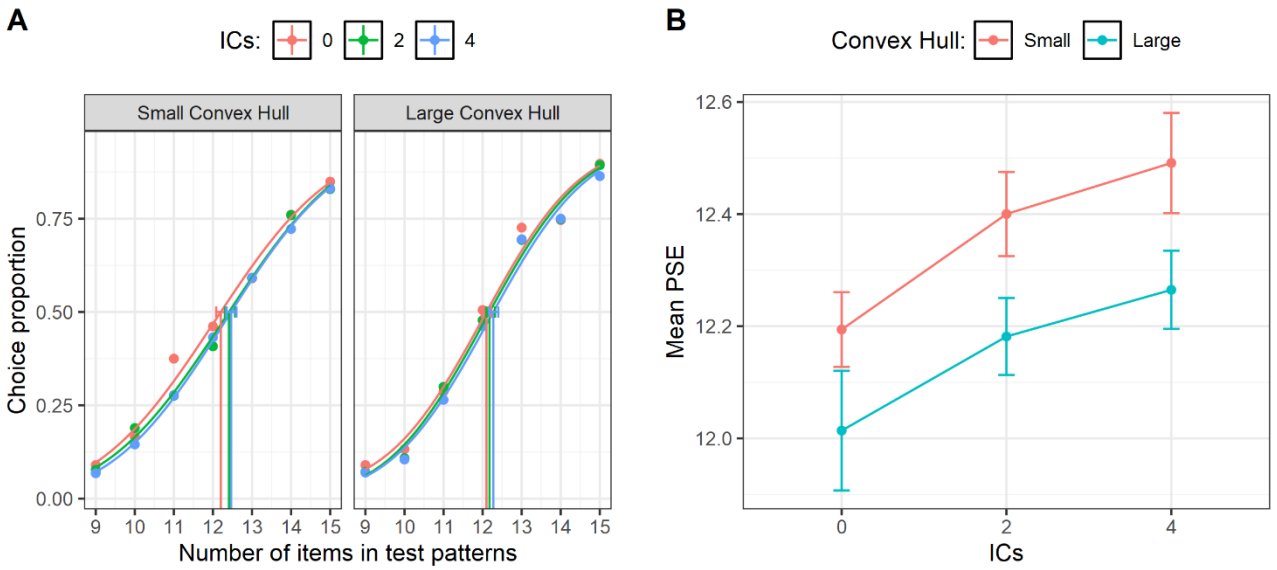


Figure 3.9: (A) Psychometric functions for 0-2-4 ICs in function of the large or small convex hull size condition, obtained pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs in function of the two convex hull size conditions and y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

3.5.3. Discussion of Experiment 3

The results of the Experiment 3 show that PSE was greater when the convex hull of test stimuli had a relatively smaller size, compared to when convex hull was relatively larger. This suggests, in line with previous evidence, that the numerosity was underestimated when the items were enclosed in a relatively smaller convex hull (e.g., Bertamini, Zito, Scott-Samuel, & Hulleman, 2016; Gebuis & Reynvoet, 2012a, 2012b) compared to a larger convex hull. Furthermore, we also found that the precision (indexed by the CoV) of the numerical representation decreased when the

convex hull size decreased. This may depend on the fact that individual items were less discriminable and were jumbled together with surrounding items when they were densely packed, leading to the underestimation effect (e.g., crowding). Alternatively, item proximity may have caused the observed effects (see Chakravarthi & Bertamini, 2020). Crucially, we also found a significant main effect of the ICs manipulation and a stable CoV across ICs conditions, thus in analogy with results from Experiment 1. Indeed, numerosity was underestimated as we increased the number of ICs. That is, within each level of the convex hull manipulation, perceived numerosity still decreases proportionally to the number of connections.

In sum, these results indicate that the manipulation of the continuous variables affects the comparison task, but numerosity processing itself is not a mere by-product of continuous variables. Indeed, a segmentation mechanism, mediated by the grouping driven by the ICs connecting lines, seems to persist even though continuous variables are experimentally manipulated as well as when they are equalized across ICs conditions.

On these grounds, in Experiment 4, we aimed at further probing whether the observed effects of ICs can be generalized to another numerical task. We therefore asked participants to perform a numerical estimation task, which requires more precise numerosity processing than the comparison task: by avoiding the occurrence of any binary-choice response mechanisms, the judgment must be accomplished on a single stimulus (Gebuis & Reynvoet, 2012b).

3.6. Experiment 4: Estimation Task with Open Inducers

3.6.1. Materials and methods

3.6.1.1. Participants

As in previous Experiments, the a priori sample size analysis suggested that a minimum of 16 participants was required to find a moderate to high effect size for the main effect of ICs.

A total of 20 undergraduate and postgraduate students (mean age ± sd = 23.45 ± 5.01 years, 13 females, 17 right-handed), were recruited for Experiment 4. They all had normal or correct-to-normal vision and they were unaware of the hypothesis of the experiment.

3.6.1.2. Stimuli and Procedure

For this experiment, we employed the test stimuli of the Experiment 1. Participants were required to explicitly estimate the number of objects in the stimuli. Task instructions were provided orally by the experimenter and were also presented on the display. No information was given regarding the range of numerosity or the presence of illusions. Each trial began with a black background (RGB = -1, -1, -1) lasting 1000 ms, followed by a central fixation cross (Font: Times; Size: 16 pixels; RGB = 0, 0, 0) which was kept on the screen for 1000 ms, and then a stimulus appeared centrally for 400 ms. After the offset of the central stimulus, a message (*"Estimation:"*; Font: Arial; Size: 30 pixels; RGB = 1, 1, 1) appeared centrally on the screen signaling that estimation was allowed. Then, the subject had to type the answer over a numerical keypad of a standard keyboard and had to press the space bar to confirm the response (Figure 3.10). The typed answer appeared on-line on the screen during the digitation and subjects were allowed to modify their subjective estimation before confirming the input. After the space bar was pressed another trial began. The experimental phase was preceded by a brief training composed of 12 trials (without feedback) in which only the stimuli with 9, 11, 13, 15 items and 0, 2 or 4 ICs were projected. After the training phase, two counterbalanced blocks of 168 trials (336 total trials) presented in random order were projected. The blocks were separated by a forced pause of 3 minutes and the whole experiment took approximately 45 minutes.

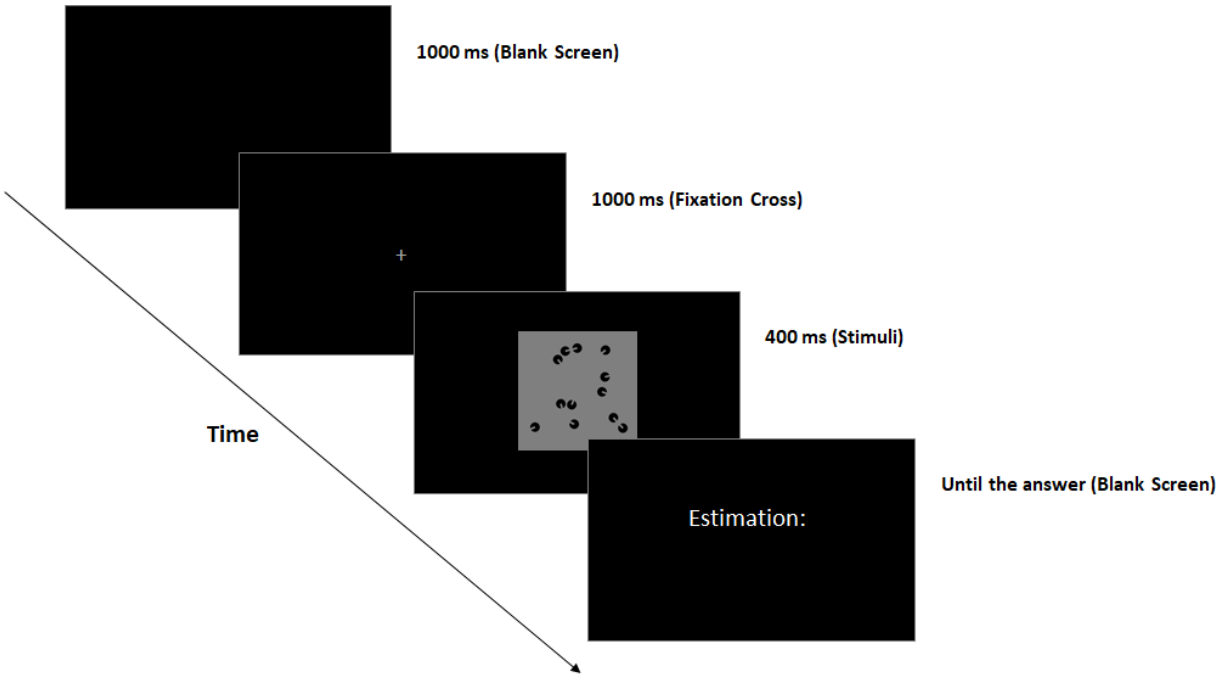


Figure 3.10: The estimation task. Subjects had to estimate the number of objects contained in the set, by typing the answer on the numerical keypad of a normal pc keyboard.

3.6.2. Results

Data were analyzed with R-Studio and Jamovi. Before running the analysis, we inspected the data for the presence of possible outliers. From the initial sample of 6720 observations, we deleted 11 empty observations (the subjects pressed the space bar without typing any answer). We also eliminated the observations that were 2 standard deviations below or above the grand mean (13 observations; < 1% of the data). Next, we performed a 3 x 7 repeated measures ANOVA with the number of ICs (0, 2, or 4) and the target numerosity (9, 10, 11, 12, 13, 14, 15) as within-subject factors and the mean subjective estimations as dependent variable. The Greenhouse-Geisser epsilon (ε) correction for violation of sphericity was applied when needed and original *F*, *df* and corrected *p*-values were reported.

The main effect of numerosity was significant, $F(6, 114) = 235.04$, $\varepsilon = .26$, $p < .001$, $\eta^2_p = .92$, showing that mean estimates increased with target numerosity (Figure 3.11A). Post-hoc paired t-test comparisons (Bonferroni correction) revealed that all the target numerosities were significantly different from each other (all $p < .05$). A polynomial trend analysis shows a significant linear trend, $t = 37.4$, $p < .001$, and a significant quadratic trend, $t = -2.66$, $p < .01$.

Critically, the main effect of the number of ICs was also significant, $F(2, 38) = 5.29$, $\varepsilon = .57$, $p = .03$, $\eta^2_p = .21$. As expected, subjective estimations decreased as we increased the number of connections (Figure 3.11B). Post-hoc comparisons (Bonferroni correction) revealed indeed a significant difference between the 0 ICs and 4 ICs condition, $t(38) = 3.20$, $p = .01$, while the mean estimations between 0 ICs and 2 ICs, $t(38) = 1.07$, $p = .86$, as well as between 2 ICs and 4 ICs, $t(38) = 2.12$, $p = .12$, were not statistically different from each other. Furthermore, we found a statistically significant decreasing linear trend only, $t = -3.19$, $p = .003$. Finally, there was no significant interaction (Figure 3.11C) between target numerosity and number of ICs, $F(12, 228) = 1.72$, $\varepsilon = .48$, $p = .13$, $\eta^2_p = .08$.

We also analyzed the Coefficient of Variation with an analogous 3 x 7 ANOVA. In line with the psychophysical models of numerosity representation (e.g., Whalen et al., 1999), mean estimates and response variability should both increase with the target numerosity, resulting in a constant Coefficient of Variation across the numerosity range (i.e., CoV = standard deviation of mean response/mean response). Results of the repeated measures ANOVA with numerosity and number of ICs as within factors and the CoV as dependent variable showed no significant effect of the target numerosity, $F(6, 114) = 2.31$, $\varepsilon = .63$, $p = .07$, $\eta^2_p = .10$, of the number of ICs, $F(2, 38) = 1.72$, $p = .19$,

η^2_p = .08, nor of the interaction, $F(12, 228) = 1.33$, $\varepsilon = .47$, $p = .25$, $\eta^2_p = .06$. As expected, therefore, the CoV was constant across all the experimental conditions (Figure 3.11D).

Furthermore, as in the previous experiments, we found a significant positive correlation among individual CoVs for each condition (See Supplementary Materials for more details; see also Figure S3.13 and Table S3.19). Finally, we also ran Bayesian statistics over the mean subjective estimations and the CoV, which overall confirmed the statistical patterns reported here (See Supplementary Materials more details; see also Table S3.17 and Table S3.18).

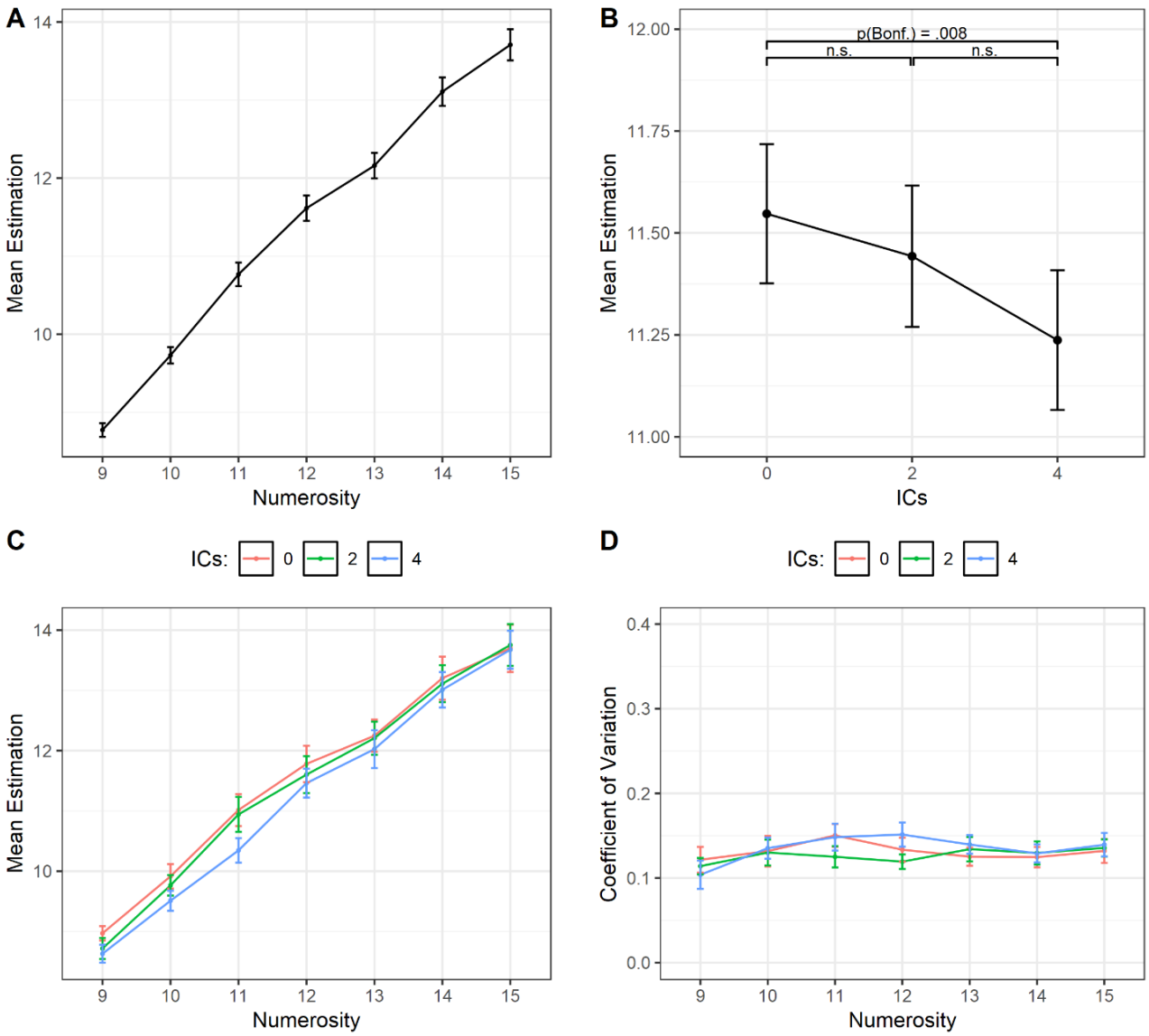


Figure 3.11: (A) Mean estimations as a function of the target Numerosity. (B) Mean estimations as a function of the number of ICs. (C) Mean estimations as a function of the target Numerosity and the number of ICs. (D) CoV as a function of the target Numerosity and of the number of ICs. The error bars represent ± 1 SEM.

3.6.3. Discussion of Experiment 4

The results of Experiment 4 suggest that the underestimation effect reported in Experiment 1 is task-independent. Indeed, we found a significant main effect of the ICs even in the estimation task. Critically, we also found that the Coefficient of Variation was stable, suggesting that subjects estimated the numerosity with an equal precision across target numerosities and ICs conditions. This pattern of results seems to confirm that numerosity representation had a scalar variability as predicted by psychophysical models, which assume that numerosities in the brain are represented as random variables with Gaussian distribution and an increasingly standard deviation proportional to the mean (Whalen et al., 1999). In such model of numerosity representation the CoV, an index of the Weber fraction (e.g., Halberda & Odic, 2014), is indeed constant across numerosities.

3.7. General Discussion

Here, we aimed at systematically probing the alleged importance of discrete mechanisms that have been firmly questioned by the theories based on surrogate features, by means of a manipulation exploiting visual illusions that allows us to control the impact of continuous variables. More specifically, by grouping individual items into pairs through ICs lines and keeping constant all low-level features of the stimuli across conditions, we investigated whether non-symbolic numerosity processing depends on the number of visually segmented entities in the scene (e.g., discrete numerosity representation) or rather over summary statistics computed from low-level features (e.g., total surface, density, etc.) in the raw visual input.

Our findings provide strong evidence for the role of visual segmentation mechanisms in driving numerosity perception. Indeed, by means of a comparison task, in Experiment 1 we showed that illusory connections lead to a numerical underestimation effect, as more items need to be present in the visual stimulus in order to be judged as numerically equal to the reference under illusory condition. Importantly, the specificity of our manipulation was further corroborated in Experiments 2A and 2B, in which we showed that –under similar experimental conditions– the underestimation effect disappears in the absence of Kanizsa-like ICs lines (i.e., as tested through closed inducers, with both 4-pixels and 1-pixel closing line thick). Furthermore, in the Experiment 3 we found that both discrete and continuous features could be independently processed during a numerical comparison task. Indeed, numerosity was underestimated when we increased the number of ICs, as in previous experiments, but also when the convex hull size of test stimuli was relatively smaller. Finally, these findings were extended to an explicit estimation task (Experiment 4): in this case, the subjective estimations decreased as we increased the number of connected pairs in the stimuli, suggesting that the connectedness effect might be task-independent.

These findings are in line with previous studies suggesting that the illusory connections would determine a mandatory binding of the items into unified objects due to the so-called element connectedness principle (Palmer & Rock, 1994), in turn indicating that numerosity estimation would operate over discrete segmented inputs rather than over the raw low-level (i.e., “continuous”) features in the visual input. However, the previous studies that used physical connections for the grouping manipulation (Franconeri et al., 2009; He et al., 2009; He et al., 2015) did not control whether the underestimation might have been induced by the task-irrelevant lines responsible for reducing the visibility of the task-relevant items in the pair. The only exception is represented by

the study of Kirjakovski and Matsumoto (2016), in which ICs lines were used to avoid this potential confound. Yet, all these previous investigations did not control for the spatial profiles of the stimuli (e.g., convex hull, density, occupancy index), therefore leaving open the possibility that some low-level features may have affected the reported findings.

In striking contrast, in the current study we carefully controlled all the possible methodological confounds (e.g., occupancy index/convex hull). In particular, we strictly controlled the spatial distribution of the items across ICs conditions, by having sets with the same continuous variables across all illusory conditions. Our results, therefore, suggest that the perceived numerosity of a numerical set is reduced when we increase the illusory connections, even though the item distribution and the spatial profile are kept constant across the different levels of connectedness. In other words, since our test stimuli have the same item spacing, the same total contour (e.g., high spatial frequency) and the same object size and convex hull (e.g., low spatial frequency) across the levels of connectedness, our findings cannot be explained by current models of the ANS tapping on continuous variables (e.g., Allik & Tuulmets, 1991; Dakin et al., 2011; Durgin, 2008; Morgan et al., 2014). In fact, a model that predicts a numerosity representation based on the weighting of all the single visual cues would have predict no differences in estimations, as the weights were kept constant across conditions (e.g., Gebuis et al., 2016). Rather, our data indicate that number processing may operate over perceived discrete (perceptual) objects in the set, and not (only or mainly) on raw low-level features or continuous variables.

These findings are perhaps not that surprising if considered in the context of Gestalt theories, which describe the different rules that the visual system applies to combine various visual fragments into a coherent object separated from its background and other objects. Connectedness is certainly one of such powerful rules, as it is crucial for early figure/ground segmentation and objecthood perception (e.g., Palmer & Rock, 1994). Accordingly, it has been suggested that the visual system decomposes the optical scene into coherent “proto-objects”, using intrinsic visual properties of the parts, like orientation, collinearity, closure, curvature, color, and shadows (Driver, Davis, Russell, Turatto, & Freeman, 2001; Rensink, 2000). Furthermore, it has been suggested that there are multiple hierarchical levels of organization within a visual scene, such as image- and scene-level representations (e.g., Enns & Rensink, 1990). In this sense, numerosity representation may as well arise at a scene-level of visual processing. One possibility is that visual numerosity processing might be based on the discrete number of items, but that an “item” can be represented by the perceptual group formed at the highest level of the observers’ hierarchical representation of the visual image.

The collinear pairs of inducers would be treated as unified input-units (e.g., scene-level representation) for the numerical computation, rather than being treated as separate elements (e.g., image-level representation), corresponding with the subjective perceptual experience of the observer. In fact, similar types of global biases have been also reported for other hierarchically organized objects, such as (global) letters composed of other smaller (local) letters (Navon, 1977).

In addition, some studies have shown that connectedness principles can affect the processing of visual information and can enhance the quantity of information stored in visual working memory, by linking image structures into hierarchical chunks (e.g., Xu, 2006, 2008). On these grounds, we hypothesize that the visual system may compute a numerosity representation through the identification of the segmented objects, processing the connected units as one whole object.

Interestingly, connectedness and other Gestalt rules, which have been largely used to study the phenomenology of visual perception, have neurobiological bases (e.g., Roelfsema & Houtkamp, 2011). This view finds support from recent neuroimaging evidence: for instance, in the study by Fornaciai and Park (2018), participants had to passively view arrays of connected (i.e., though with physical lines) or unconnected dots, while they underwent functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) recordings. The effect of connectedness was found 150 ms after the onset of stimuli in the EEG session and in third visual area (V3) in the fMRI session, but not in earlier neural activity (Fornaciai & Park, 2018). This suggests that segmentation of the stimulus into perceptual units as input for the numerosity representation occurs at around 150 ms post-stimulus onset in V3 cortex (for complementary evidence see, He et al. 2015; DeWind et al., 2019). Hence, these findings suggest that a coarse numerosity processing can be found in early visual cortex (e.g., V3), after the accomplishment of some previous perceptual-organization process on the raw visual input.

We, therefore, agree with those proposals suggesting that the foundation of numerosity perception might be rooted in the evolved ability to separate figures from ground and from neighbor objects (e.g., Fornaciai & Park, 2018; Franconeri et al., 2009; Kirjakovski & Matsumoto, 2016). These fundamental visual-segmentation mechanisms have been found not only in humans and monkeys (Peterhans & von der Heydt, 1989, 1991) but also in fishes (e.g., Sovrano & Bisazza, 2008, 2009), and chicks (e.g., Regolin & Vallortigara, 1995), suggesting that they could represent the visual substrates that may support non-symbolic numerical processing both in primates and non-primate species. For example, animal camouflage is based on the strategy to minimize the number of visual

cues that differentiate an organism from the environment: as a consequence, the ability to identify illusory contours may have evolved as anti-camouflage mechanism (Ramachandran, 1987) to detect partly occluded or fragmented objects in the visual scene using collinearities of the parts to reconstruct the "whole" object, even when object information is substantially missing.

A final speculation is that the visual system may have recycled neural networks originally evolved for object detection for a new advantageous purpose, like keeping track of the approximate number of segmented objects in the visual scene (e.g., Anderson, 2010). This speculation seems to be partially confirmed by recent computational models (DeWind, 2019; Nasr, Viswanathan, & Nieder, 2019), which suggest that the ANS may have naturally emerged from basic visual circuits originally dedicated to object perception.

In sum, we suggest that a discrete numerosity representation might be affected by Gestalt principles of perceptual organization that govern the figure/ground segmentation and perceptual units formation. In other words, the structuring of the sensory scene (perceptual organization), rather than the simple raw amount of low-level features, might shape the final number of coherent objects and the subjective "numerical magnitude" that we perceive. We pinpoint that our theoretical view is not incompatible with Dehaene's model, which suggests a "normalization" of the visual features before numerosity extraction. The fact that "dumbbells" (connected) items are counted as single input units independently of their shape, should suggest that a sort of size normalization would be at play. At the same time, we do not exclude the role of continuous features during numerosity processing (as we also showed in the Experiment 3), but rather, we empirically showed that –at least in some situations– they cannot explain human behavioral performance.

In line with this possibility, it has been shown that both continuous physical dimensions (e.g., item size or cumulative surface) and discrete number information may be automatically extracted in Stroop-like tasks even when they are irrelevant to the task (e.g., Hurewitz et al., 2006; Nys & Content, 2012), and they may interact or compete for behavioral control, perhaps in a late decisional stage (Franconeri et al., 2009; Leibovich & Henik, 2014). A final intriguing possibility is that contextual factors may promote different strategies based on discrete numerosity or low-level cues (e.g., Dietrich, Nuerk, Klein, Moeller, & Huber, 2019). It is not unlikely that different systems may be recruited for magnitude processing, but perhaps contextual factors (e.g., range or the number of items in the scene) or task constraints (e.g., stimulus exposure time) may be responsible for activating one of these systems or for promoting different strategies based on the discrete numerosity or rather on the low-level visual cues (e.g., Dietrich, Nuerk, Klein, Moeller, & Huber,

2019). For example, it is possible that texture-density models of numerosity processing (Dakin et al., 2011; Durgin, 2008) are well suited to explain participants' performance under specific visual contexts, such as highly cluttered scenes (Anobile et al., 2014, 2016), or during ultra-rapid stimulus presentation, where power spectrum content may provide sufficient information to have a "gist" of the scene using rough statistical summary of the image (e.g., spatial frequencies or luminance), without necessarily recurring to the segmentation of the single objects in the scene (e.g., Oliva & Torralba, 2006; Utochkin, 2015)

Furthermore, it has been shown that the weight of continuous sensory cues can be modulated by changes in task instructions, task difficulty and stimulus duration (Leibovich, Henik, & Salti, 2015; Leibovich-Raveh, Stein, Henik, & Salti, 2018). Thus, it seems that the weight of the sensory cues affecting number processing is not static but could be modified in a flexible way by the manipulation of bottom-up and top-down factors, as well as their interaction. Another contextual factor that has been shown to affect the weight of sensory cues is the trial-history. Indeed, it has been demonstrated that performance is enhanced when subjects are presented with easier ratios in the first trials compared to when presented with the hardest ratio first (Odic, Hock, & Halberda, 2014). Finally, some studies showed that the *perceived* weight of continuous features affects the performance, by manipulating the actual visual context in which numerical stimuli were embedded by means of classic Gestalt size illusions such as the Ebbinghaus (Picon et al., 2019) and the Müller-Lyer illusion (Dormal, Larigaldie, Lefèvre, Pesenti, & Andres, 2018).

It follows that future models of numerical encoding will need to reconcile these competing theoretical views, considering more extended approaches for parallel processing of discrete numerosity and continuous signals.

3.8. Conclusions

In conclusion, the present study contributes to unveil how our mind extracts numerosity from the visual scene making sense of its complexity. In particular, our work indicates that the visual system can extract a numerosity representation from unified "discrete" surfaces, even when object information is incomplete at the input level. This study provides, therefore, strong evidence to the theoretical role of segmentation visual processes in number perception even though continuous features are still processed. We suggest that natural visuo-biological constraints might have shaped

the development of a neural system dedicated to “number processing”, which would be partially independent from raw low-level cues and rather based on discrete segmented visual entities.

Chapter 4

4.1. Non-symbolic numerosity in sets with illusory contours exploits a context-sensitive, but contrast-insensitive, visual boundary formation process.

Chapter adapted from: Adriano, A., Rinaldi, L., & Girelli, L. (2021). *Attention, Perception, & Psychophysics*, 1–16. https://doi.org/10.3758/s13414-021-02378-y

4.2. Introduction

Recently, visual illusions have been successfully employed to understand which visual features represent the building block of numerosity perception, because they can be used to selectively manipulate particular features of numerical sets without altering other physical features in the image (e.g., Dormal, Larigaldie, Lefèvre, Pesenti, & Andres, 2018; Picon, Dramkin, & Odic, 2019; Pecunioso, Petrazzini, & Agrillo, 2020). For instance, the so-called connectedness illusion has been adopted to manipulate the perceived segmentation (or grouping strength) of the items in the set, keeping constant the low-level features across connectedness levels (Franconeri et al., 2009; He et al., 2009). In these studies, irrelevant lines were used to connect and manipulate the number of dot-pairs. This manipulation proportionally reduced the perceived numerosity likely because the visual system was forced to process two dots as a single unified perceptual object, as suggested by the grouping principle of element connectedness (Palmer & Rock, 1994), representing hence the input-units of visual numerical computation. These findings were also recently replicated with grouping manipulations in which Kanizsa-like illusory contour (ICs) lines were used instead of actual physical lines, as the latter may obscure the task-relevant items (e.g., Adriano, Rinaldi, & Girelli, 2021; Kirjakovski & Matsumoto, 2016). ICs are visual experiences of objects whose edges are not defined by physical luminance discontinuities with the background (Nieder, 2002; Wagemans et al., 2012), and thus are well suited to replace the physical lines to simulate connections. In short, studies manipulating either the real or the illusory connecting lines strongly suggest that non-symbolic numerosity would be extracted from discrete segmented objects rather than from raw low-level features of an unsegmented scene. However, in studies manipulating the ICs connections (Adriano,

Rinaldi, & Girelli, 2021; Kirjakovski & Matsumoto, 2016), the illusory lines generated by the inducers were accompanied by a subjective brightness enhancement (e.g., since inducers were darker than the background), which can represent a further "perceived" continuous confound occurring when illusory lines in the sets were increased. The perceived amount of continuous cues in numerical stimuli (rather than mere physical information), as manipulated with classic size illusions, indeed may affect numerical tasks (Dormal et al., 2018; Picon et al., 2019). Furthermore, it has been shown that pupillary diameter may decrease with the illusory brightness enhancement induced by classic Kanizsa illusion and similar brightness illusions (Laeng & Endestad, 2012; Zavagno, Tommasi, & Laeng, 2017), which may reduce the light information sampled by the eyes and the visual input, in turn explaining the underestimation effect. Indeed, according to an influential neural theory of vision, ICs emerge as the result of the synergy between two separated but complementary streams in early visual cortex (also known as Form-And-Color-And-Depth model or FACADE, Grossberg, 2014): the boundary completion system and the surface filling-in system. Consequently, there is often a *subjective* or *perceived* change in brightness that human observers perceive at such illusory contours, but the two processes are governed by two complementary mechanisms. According to this model, one important characteristic separating the two systems is that boundary completion pools across opposite contrast polarities, and thus occurs in a manner that is *insensitive* to contrast polarity of inducers, whereas surface filling-in does not pool opposite contrast polarities and is *sensitive* to contrast polarity creating percepts of brightness and color. This "two-systems" division is reinforced by psychophysical studies showing that observers did not perceive strong subjective brightness difference between the illusory surface and the (gray) background when inducing elements had opposite contrast signs (e.g., two white and two black inducers, as in the reverse-contrast Kanizsa square illusion). That is, the perceived brightness of illusory surface may strongly diminish, or even disappear, when the inducing elements have opposite contrast signs (Grossberg, 2014). In this case, the local signals of differential brightness generated by the individual inducer features should cancel each other out: the brightness induction due to the black-to-gray pac-man inducers should balance the darkness induction due to the white-to-gray pac-man inducers. Consequently, no global representation of a brightness difference can be extracted from the stimulus. By contrast, the illusory boundary processing may not be affected by variations in the contrast polarity of the inducing elements since the illusory figure is still perceived by the observers when inducers have opposite contrast sign (e.g., Dresp et al., 1996; Grossberg, 2014; Matthews & Welch, 1997).

Taking advantage of these studies, the current work was aimed to disentangle which of the two processes (i.e., boundary completion or surface filling-in) actually drive the underestimation effect triggered by Kanizsa-like ICs lines (Kirjakovski & Matsumoto, 2016). Hence, we carried out three experiments in which we modulated the number of aligned inducers triggering ICs (0, 2 or 4 connecting lines) as a function of the contrast polarity (e.g., positive or negative with respect to the background) of the inducing elements. In Experiment 1, inducer-pairs triggering the ICs were formed by light-to-gray open inducers (all white) or dark-to-gray open inducers only (all black). In Experiment 2, inducer-pairs triggering the ICs had always opposite contrast polarity (one black and one white) compared to the background. In Experiment 3, to exclude further confounds due to item orientation statistics, aligned inducers had opposite contrast polarity but were closed with a thin line.

If the underestimation of test stimuli is merely due to the perceived brightness enhancement (black inducers triggering ICs brighter than ground), and hence to the surface fill-in system, the effect should be reversed when ICs are darker than ground (white inducers, Experiment 1) and, crucially, no underestimation should be found when inducers of opposite contrast polarity (which suppresses brightness enhancement) were aligned triggering the ICs (Experiment 2). On the other hand, if the underestimation effect will be preserved despite inducers with different contrast polarity (Experiment 1) or simultaneous opposite contrast polarity (Experiment 2) are aligned to create ICs-connections, we should conclude that this effect is not related to the ICs brightness itself. Finally, we predict that in Experiment 3 no underestimation should be found when illusory contours formation was prevented. In such a scenario, the boundary completion system would be the ideal candidate for explaining the observed underestimation effect (Grossberg, 2014).

4.3. Experiment 1: Single contrast polarity open inducers

In Experiment 1 we tested whether the underestimation effect reported in previous studies may depend on the *perceived* change in brightness of the generated illusory surface. We manipulated the number of pairs of aligned inducers or ICs (0, 2 or 4) and the direction of inducers contrast (e.g., positive or negative) compared to the background (inducers were drawn only in black or white over a middle gray background). Previous studies (Adriano et al., 2021; Kirjakovski & Matsumoto, 2016) employed "brighter" ICs only (generated by black inducers on a gray background), causing a rightward shift of the psychometric functions (e.g., increasing PSE) when ICs were increased. We predicted that if this pattern is merely due to perceived brightness, its direction should be reversed (e.g., leftward shift or decreasing PSE), or at least differently modulated, when darker ICs (generated by the white inducers) were presented, resulting in a significant interaction between the two independent variables (i.e., numbers of ICs and color of the inducers). Otherwise, a null interaction between ICs number and color of the inducers should suggest that numerosity underestimation driven by ICs does not depend on the illusory brightness itself (e.g., no difference in underestimation between black or white inducers).

4.3.1. Materials and methods

4.3.1.1. Participants

Because of the Coronavirus pandemic restrictions in Italy, the participants were recruited through Pavlovia, a repository and launch platform allowing online PsychoPy experiments (www.pavlovia.org). A sample of 28 participants (21 females, 7 males) took part in the study[2]. The mean age was 28.5 years (*SD* = 7.12). Handedness was assessed by asking participants which hand they typically used for writing: a total of 23 subjects were classified as right-handed. All participants had normal or correct-to-normal vision and were naïve about the purpose of the experiment. Each subject signed an on-line informed consent document before the experiment began and the study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Local Ethical Committee (protocol N° RM-2020-230).

[2] Note that we collected overall data from 34 participants, but 6 of them were discarded from the final sample prior the analysis because presented a numerical acuity (e.g., coefficient of variation) that fell above or below the interquartile range (e.g., ± 2 SD) of the distribution in one or more conditions, suggesting a very poor or random performance.

4.3.1.2. Stimuli

The stimuli set were generated off-line by a custom Python/Psychopy script and projected by means of an on-line Psychopy routine (Peirce, 2007). Stimuli were constructed with the same specifications as in Adriano, Rinaldi and Girelli (2021; *see Chapter 3*), but with the suited changes to manipulate the contrast polarity of inducing elements (black or white, see Figure 4.1). The whole experimental set was composed of 168 test stimuli (56 random spatial patterns cloned across the 3 levels of connection, half drawn with black inducers and half with white inducers) and 168 reference stimuli (56 random spatial patterns, 28 with all black inducers and 28 with all white inducers, repeated within the 3 levels of connection) generated off-line.

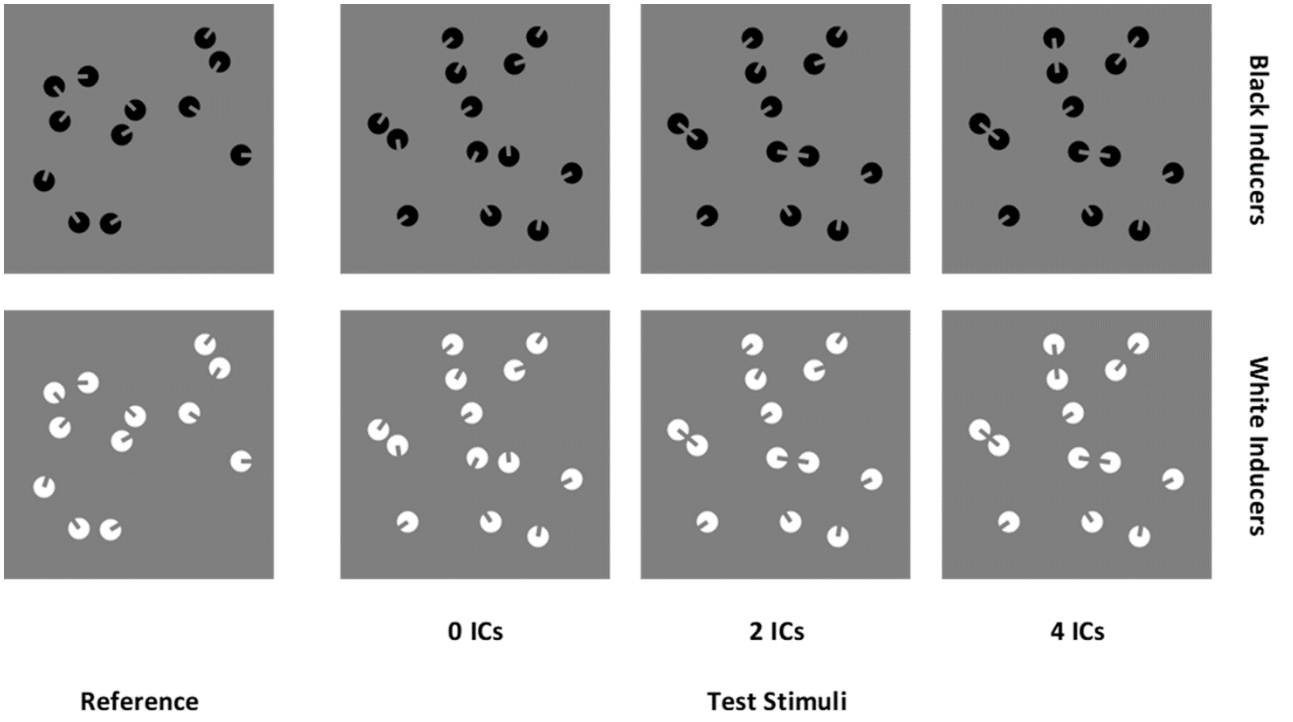


Figure 4.1: An example of stimuli pairs used in Experiment 1. The reference stimulus was always composed of 12 items. Test patterns varied from 9 to 15 items and contained 0 ICs, 2 ICs or 4 ICs. All the test stimuli had the same convex hull, density and total surface across the conditions. Half of the stimuli were drawn with white inducers and the other half with black inducers.

4.3.1.3. Procedure

The procedure follows the Experiment 1 in *Chapter 3*. Two counterbalanced blocks of 336 randomly ordered trials were presented, for a total of 672 experimental trials (16 trials for each of the 42 conditions), separated by a self-paced pause at the half of the experiment. Hence, all the

experimental manipulations were within-blocks (50% of the trials with all black or all white inducers) and thus within-subjects (Figure 4.2). The whole experiment lasted around 40-45 min.

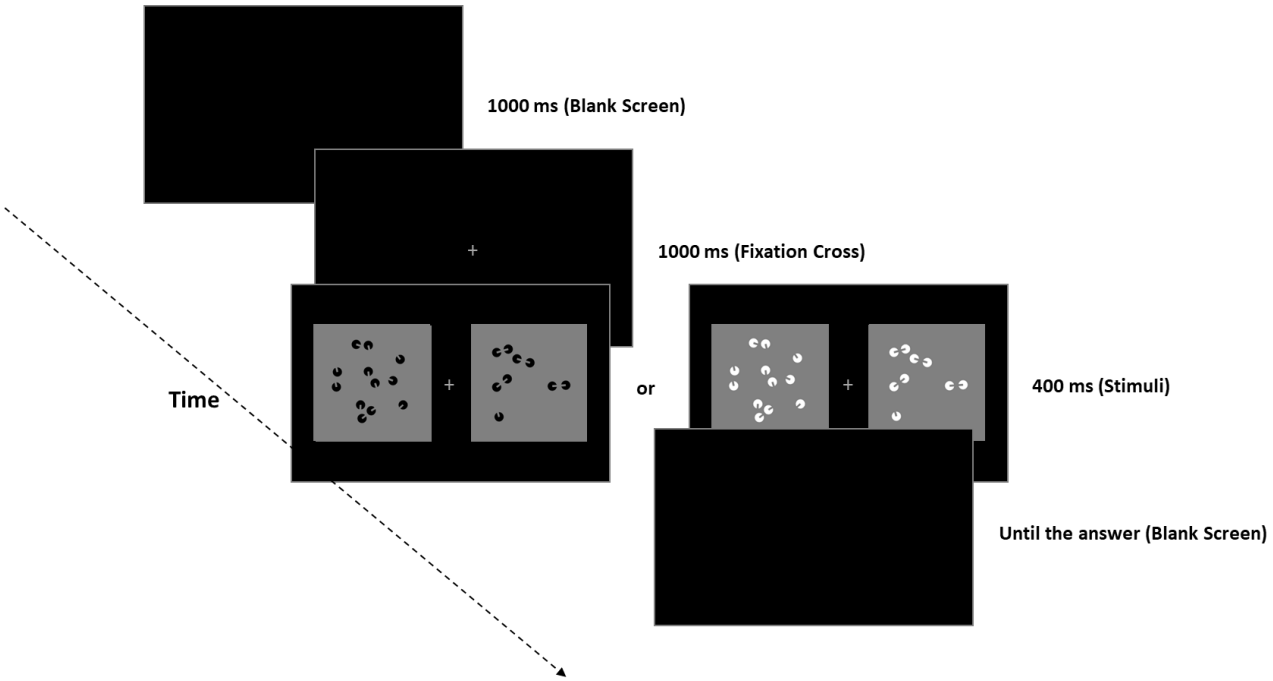


Figure 4.2: The discrimination task. Subjects had to decide which stimulus was numerically larger by typing the relative key to specify the left or right stimulus side (F or J key). The side of reference and test pattern was balanced and randomized. In half of the trials, we presented white stimuli (test and reference patterns), while in the other half black stimuli.

4.3.1.4. Data analysis

The data were analyzed with *R-Studio* (2018, v. 3.6.2; http://www.rstudio.com/) and *Jamovi* (2019, v. 1.1.5; https://www.jamovi.org) softwares as in the Experiment 1, *Chapter 3*.

Two separated two-way repeated measures ANOVAs were performed with the number of ICs (0, 2 or 4) and the inducers color (white or black) as within-subjects factor and with the mean PSE or the mean CoV as dependent variables, respectively. The Greenhouse-Geisser epsilon (ε) correction for violation of sphericity was applied when needed, and original *F*, *df* and corrected *p*-values were reported. We also run Bayesian analyses (see Supplementary Materials).

4.3.2. Results

As can be visually observed in the figure plotting the psychometric functions obtained pooling over the aggregate data of the whole sample (Figure 4.3A), we found a rightward shift of the psychometric curves when darker or brighter ICs were increased, suggesting a systematic underestimation of the perceived numerosity as we increased the number of connections regardless of the color of the inducers (note that this graph is reported to illustrate the technic, but all subsequent analysis was done with similar functions over individual subjects; see Figure S4.1 for individual data). The analysis of individual PSEs showed a significant main effect of the number of ICs, $F(2, 54) = 8.06$, $p < .001$, $\eta^2_p = .23$. That is, the PSEs increased with the number of ICs (Figure 4.3B). Such a pattern, was confirmed statistically by means of post-hoc comparisons (Bonferroni correction), revealing a significant difference between 0 ICs and 4 ICs, $t(54) = -3.98$, $p = < .001$, while no significant difference was found between 0 ICs and 2 ICs, $t(54) = -1.55$, $p = .37$, and between 2 ICs and 4 ICs, $t(54) = -2.42$, $p = .056$. A polynomial trend analysis showed a significant linear trend only, $t(54) = 3.98$, $p < .001$. Furthermore, the main effect of inducers color was not statistically significant, $F(1, 27) = 1.83$, $p = .187$, $\eta^2_p = .06$, and crucially no significant interaction between the two factors was found, $F(2, 54) = .57$, $\varepsilon = .80$, $p = .52$, $\eta^2_p = .02$, suggesting a similar underestimation with black or white inducers. Furthermore, the analysis of the CoV of the psychometric functions revealed no significant main effect of number of ICs, $F(2, 54) = 2.1$, $\varepsilon = .80$, $p = .142$, $\eta^2_p = .07$, no significant main effect of inducers color, $F(1, 27) = .43$, $p = .51$, $\eta^2_p = .016$, and no significant interaction, $F(2, 54) = 1.59$, $p = .21$, $\eta^2_p = .056$, suggesting that participants numerical acuity was stable across all the conditions as predicted by the Weber's law (see Supplementary Materials and Figure S4.2). In addition to frequentist analyses, we also ran Bayesian ANOVAs over both the PSEs and the CoVs with the number of ICs and inducer colors as independent variables. These additional analyses confirmed the main results reported here (see Supplementary Materials for more details; see also Table S4.1 and Table S4.2).

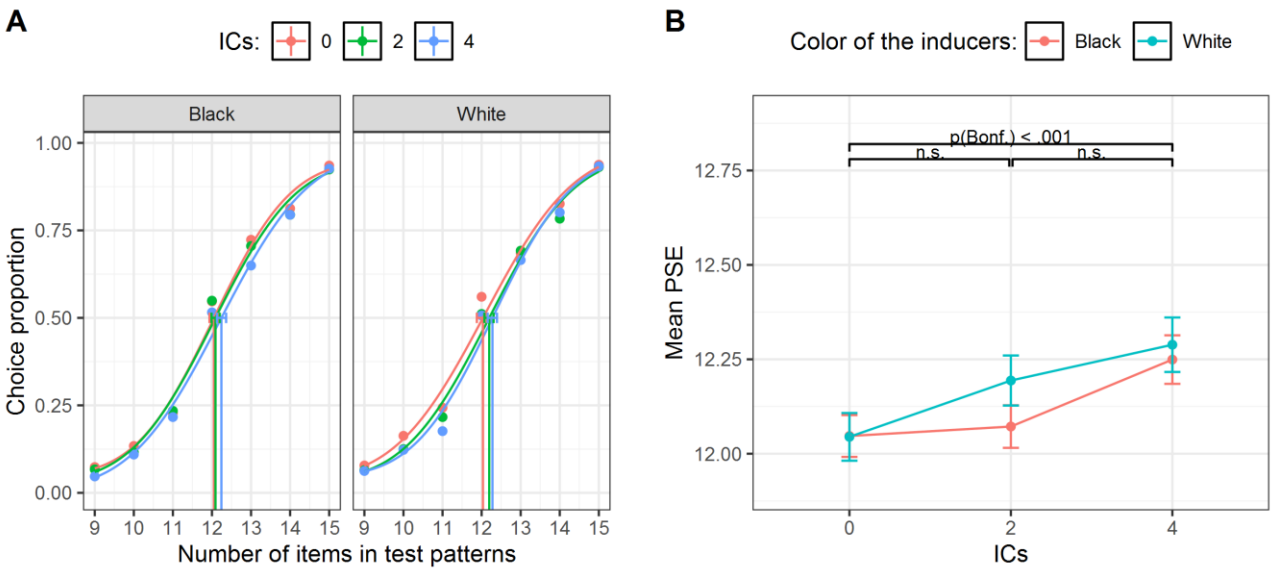


Figure 4.3: (A) Psychometric functions obtained in Experiment 1 for 0-2-4 ICs as a function of the color of the inducers (black or white), obtained pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs as a function of the color of the inducers, and the y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

4.3.3. Discussion of Experiment 1

The results of Experiment 1 revealed that the color of the inducers did not affect participants' numerical estimations. Furthermore, the effect of ICs was found to be statistically significant, and crucially no interaction was found with the inducer color type, showing a similar increasing pattern in PSE over both color type. This suggests that more items were required in test patterns to be perceived numerically equal to reference when inducer pairs were aligned, independently of the polarity (or contrast direction) of the inducers and the relative change in surface brightness (brighter or darker than the background). Hence, it is unlikely that filling-in process might be the source of this effect, indirectly suggesting that the boundary system (which is insensitive to contrast polarity) would drive the underestimation effect. To directly probe this possibility, in the Experiment 2, we tested participants with stimuli composed of opposite-contrast inducers, which should strongly reduce the perceived brightness of the generated ICs shapes (Grossberg, 2014).

4.4. Experiment 2: Reverse-contrast polarity open inducers

To corroborate and extend the results of the first experiment, in the Experiment 2 we tested whether the underestimation effect is preserved when ICs-lines were generated by inducers with reverse contrast polarity. If the underestimation effect is merely a by-product of the IC brightness, no underestimation effect should be found when reverse contrast polarity inducers were aligned (e.g., one black and one white), since in this case the perceived difference in luminance of the illusory surface is strongly reduced (e.g., Dresp et al., 1996; Grossberg, 2014; Matthews & Welch, 1997). Otherwise, if the boundary contour system (which is insensitive to contrast polarity of inducers) drives the effect, the underestimation pattern should be preserved even when inducers have opposite contrast polarity.

4.4.1. Materials and methods

4.4.1.1. Participants

A new sample of 23 participants (14 females) was recruited for the second on-line study[3]. The mean age was 30.73 years (*SD* = 7.91). A total of 22 participants were classified as right-handed. All the subjects had correct or correct-to-normal vision and were naïve to the goal of the study.

4.4.1.2. Stimuli & Procedure

The stimuli were generated as in Experiment 1, but the inducers had mixed polarity. The reference patterns were composed by 12 “pac-man” like items, half of the items were drawn in white (RGB = 1, 1, 1) and the other half in black (RGB = -1, -1, -1). As in the first experiment, test patterns contained a variable numerosity (from 9 to 15 “pac-man” like items), but those containing an even numerosity (10, 12, 14) were constructed with an equal number of black and white inducers on a mid-gray background (e.g., Kogo, Drożdżewska, Zaenen, Alp, & Wagemans, 2014) whereas test stimuli with odd numerosity (9, 11, 13, 15) were counterbalanced, containing one free inducer in

[3] A total of 27 subjects were tested, but 4 subjects were discarded from the final sample because presented a numerical acuity (e.g., coefficient of variation) that fell above or below the interquartile range (e.g., ± 2 SD) of the distribution in one or more conditions, suggesting a poor or random performance. Hence analyses were run over a final sample of N=23 subjects.

excess drawn in black in half of the patterns (4 random visual patterns) and drawn in white in the other half (4 random visual patterns). Note that aligned inducers forming the ICs in test stimuli had always one black and one white inducer (Figure 4.4). All the items were drawn on a mid-grey background (RGB = 0, 0, 0).

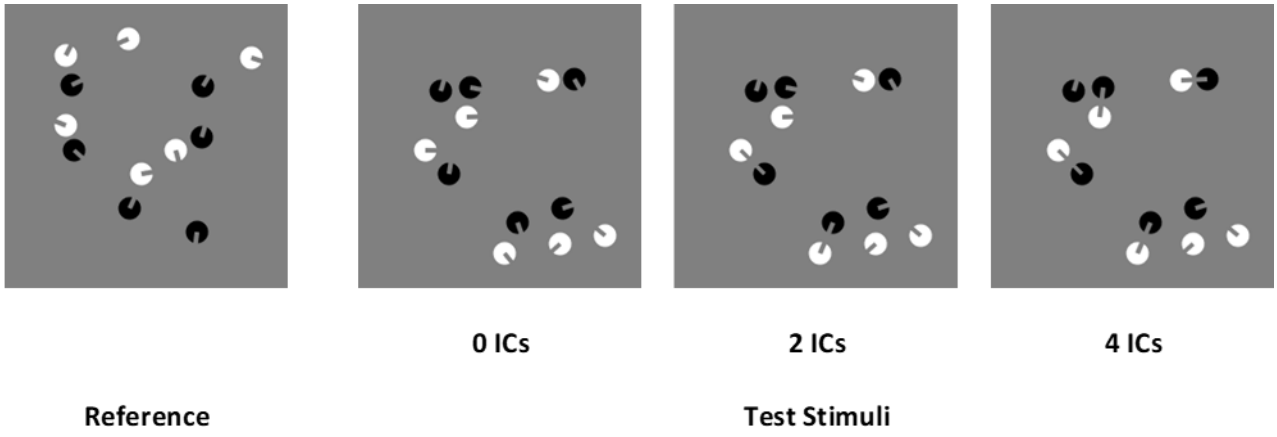


Figure 4.4: An example of stimuli pairs used in Experiment 2. The Reference set was always composed of 12 items (half black and half white inducers). Test patterns varied from 9 to 15 items and contained 0 ICs, 2 ICs or 4 ICs. All the test stimuli had the same convex hull, density and total surface across the levels of connectedness (as in the examples depicted).

4.4.2. Results

Two separate one-way repeated measures ANOVAs were carried out with either the PSE or the CoV of the psychometric functions as dependent variables and the number of ICs (0, 2 or 4) as within-subjects variable.

Visual inspection of the figure plotting the psychometric functions obtained pooling over the aggregate sample data (Figure 4.5A), suggests a rightward shift of the psychometric curves for 2 and 4 ICs conditions, thus suggesting an underestimation of the perceived numerosity as we increased the number of connections in test stimuli (see Figure S4.3 for individual data).The analysis of individual PSEs showed a significant effect of the number of ICs, $F(2, 44) = 11.14$, $p < .001$, $\eta^2_p = .33$. That is, the PSEs increased with the number of ICs (Figure 4.5B). Such a pattern, was confirmed statistically by means of post-hoc comparisons (Bonferroni correction), revealing a significant difference between 0 ICs and 4 ICs, $t(44) = -4.72$, $p < .001$, while no significant difference was found

between 0 ICs and 2 ICs, $t(44) = -2.46$, $p = .053$, and between 2 ICs and 4 ICs, $t(44) = -2.25$, $p = .087$. Furthermore, a polynomial trend analysis showed a significant linear trend only, $t(44) = 4.72$, $p < .001$.

The analysis of the CoV of the psychometric functions for the three levels of connectedness showed no significant differences across conditions, $F(2, 44) = 1.58$, $p = .21$, $\eta^2_p = .067$, suggesting an equal numerical estimation precision across ICs conditions as predicted by the Weber's law (Figure S4.4).

In addition to frequentist analyses, we also ran Bayesian ANOVAs over both the PSEs and the CoVs with the number of ICs as independent variable. These additional analyses confirmed the main results reported here (see Supplementary Materials for more details; see also Table S4.3 and Table S4.4).

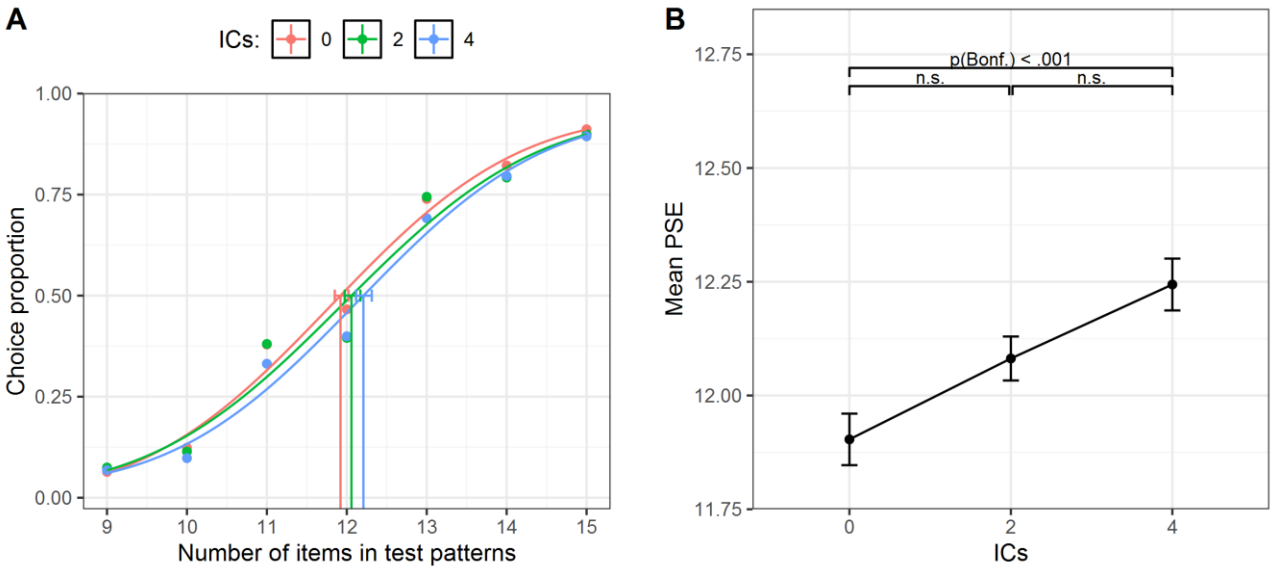


Figure 4.5: (A) Psychometric functions obtained in Experiment 2 for 0-2-4 ICs, pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs and the y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

4.4.3. Discussion of Experiment 2

The increase in the PSEs found in Experiment 2 suggests a decrease of perceived numerosity (underestimation) caused by the grouping of few individual items into pairs. Crucially, the results of this experiment corroborate the idea that the numerical underestimation effect triggered by the ICs lines is independent from the perceived brightness enhancement that accompanies Kanizsa-like illusory figures. Indeed, numerosity was underestimated (e.g., PSE increased) as the number of ICs-lines was manipulated, even though the ICs were triggered by opposite contrast polarity inducers that did not produce substantial brightness enhancement (e.g., Grossberg, 2014). This strongly suggests that the numerosity underestimation effect found in previous studies (e.g., Adriano et al., 2021; Kirjakovski & Matsumoto, 2016) was driven by the boundaries completion system (e.g., Grossberg, 2014). Hence, the numerical underestimation effect was actually due to the binding of the inducers into a dumbbell object and cannot be explained by the sensory confounds that brightness enhancement may have involuntary introduced.

However, such findings may be still explained by a general effect due to inducers edge alignment and/or item orientation statistics rather than by the boundary completion of the illusory lines (e.g., DeWind, Bonner, & Brannon, 2020). To exclude this possibility, in Experiment 3, the “pac-man” shapes were the same as in Experiment 2, but each inducer was closed with a line to prevent ICs formation.

4.5. Experiment 3: Reverse-contrast polarity closed inducers

Previous studies showed that closing the notch of inducers with a thin-line strongly reduced the formation of Kanizsa-like ICs and blocked the secondary visual cortex (V2) response (Davis & Driver, 1994; Peterhans & von der Heydt, 1989, 1991; von der Heydt, Peterhans, & Baumgartner, 1984). Here, we predicted that if the results of Experiment 1 and Experiment 2 were due to mere inducer orientation statistics, rather than to the specific completion of the ICs lines, we should find an equal increase in the PSEs when the notch of the closed inducers were spatially aligned as in the previous experiments. A lack of effect would suggest that underestimation effect was specifically driven by the illusory boundary and not to inducers orientation statistics (Kirjakovski & Matsumoto, 2016).

4.5.1. Materials and methods

4.5.1.1. Participants

A new sample of 24 undergraduate and postgraduate students (mean age ± sd = 26.87 ± 4.19 years, 17 females, 23 right-handed), with normal or correct-to-normal vision, was recruited for the third on-line study[4]. All the subjects were naïve regarding the purpose of the experiment.

4.5.1.2. Stimuli and Procedure

The design, stimuli parameters and their generation method as well as the procedure were identical to Experiment 2. The only difference in the visual stimuli patterns was that inducers were closed with a curved line (1 pixels thick), thus completing the overall circular shape of each item. Note that stimuli spatial patterns were cloned from stimuli with open inducers (Figure 4.6). For the sake of clarity, to define the name of the experimental conditions with aligned closed inducers we adopted the same labeling as in Experiment 2 (e.g., 0 ICs, 2 ICs and 4 ICs).

[4] A total of 26 subjects were tested, but 2 subjects were discarded from the final sample because presented a numerical acuity (e.g., coefficient of variation) that fell above or below the interquartile range (e.g., ± 2 SD) of the distribution in one or more conditions, suggesting a poor or random performance. Hence analyses were run over a final sample of N=24 subjects.

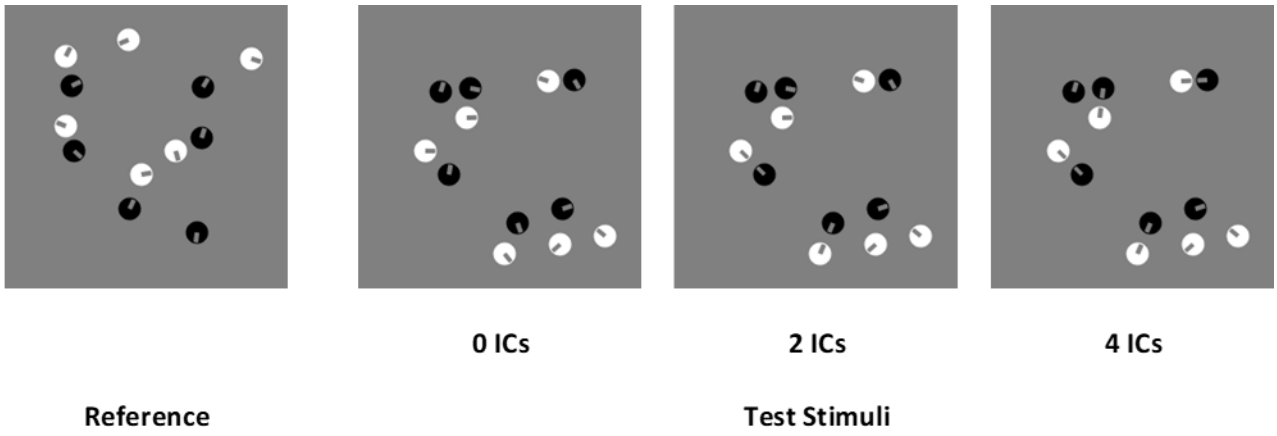


Figure 4.6: An example of stimuli pairs used in Experiment 3. The Reference set was always composed of 12 items (half black and half white inducers). Test patterns varied from 9 to 15 items and contained 0 ICs, 2 ICs or 4 ICs. All the test stimuli had the same convex hull, density and total surface across the levels of connectedness (as in the examples depicted).

4.5.2. Results

Data were analyzed as in Experiment 2 (see Figure S4.5 for the individual psychometric functions). Two separate one-way repeated measures ANOVAs were carried out with either the PSE or the CoV of the psychometric functions as dependent variables and the number of ICs (0, 2 or 4) as independent variable. As expected, the analyses showed no significant differences across the conditions, as the number of ICs did not affect the PSEs, $F(2, 46) = 1.5$, $p = .23$, $\eta^2_p = .06$, (Figure 4.7A and Figure 4.7B). Similarly, no effect of ICs condition was found for the CoV, $F(2, 46) = .154$, $p = .85$, $\eta^2_p = .007$, (Figure S4.6), suggesting an equal numerical precision across conditions. Finally, we also ran supplementary Bayesian statistics over both the PSE and the CoV (See Supplementary Materials for more details; see also Table S4.5 and Table S4.6), which overall confirmed frequentist analyses.

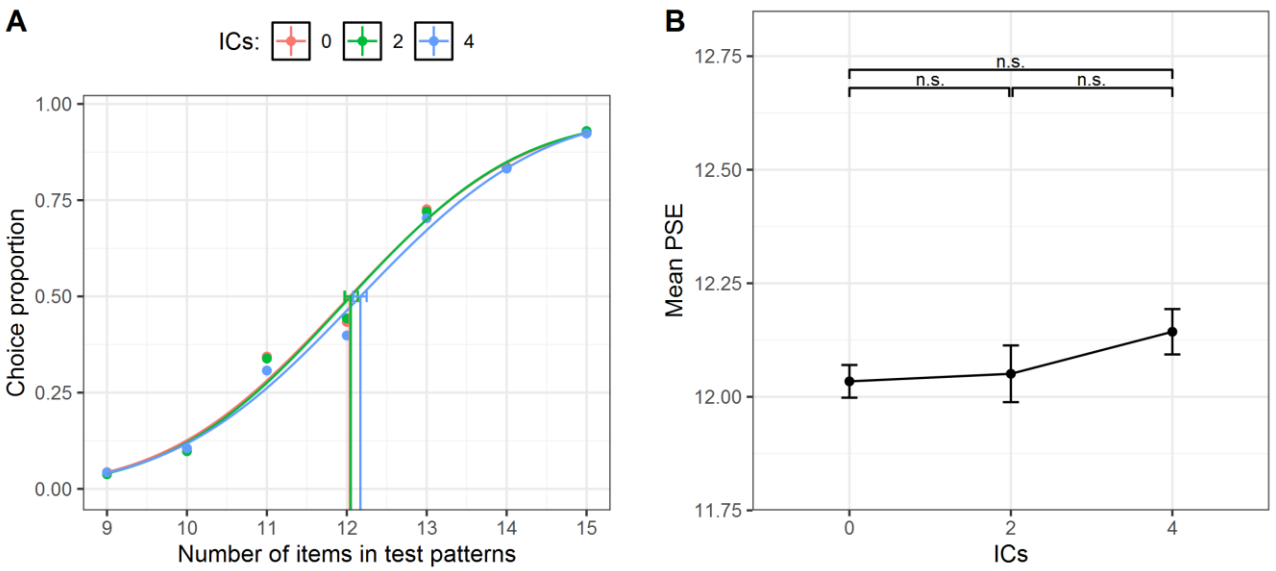


Figure 4.7: (A) Psychometric functions obtained in Experiment 3 for 0-2-4 ICs, pooling over the aggregate data of all the subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) The x-axis shows the number of ICs and the y-axis the mean PSE for each condition in the whole sample. The error bars represent ± 1 SEM.

4.5.3. Discussion of Experiment 3

The results of Experiment 3 indicate that the mere manipulation of spatial alignment (or orientation) of edges between a few inducers was not sufficient to group the items into a perceptual object producing the underestimation obtained with open inducers (e.g., Davis & Driver, 1994; Kirjakovski & Matsumoto, 2016). Furthermore, these results suggest that a mere difference in item orientation statistics across condition cannot explain the results of previous experiments (DeWind et al., 2020), replicating prior evidence with only black inducers (e.g., Adriano et al., 2021; Kirjakovski & Matsumoto, 2016). The slight increase in PSE of Experiment 3 could also be in line with previous studies using similar closed inducers (e.g., Chen, Glasauer, Müller, & Conci, 2018) but adopting a dot-localization task. Indeed, Chen and colleagues (2018) found that closing the notch does not completely suppress the effect of surface completion, especially when only a tiny closing line is used. It is therefore possible that, in Experiment 3, closed inducers still triggered a weak grouping (as showed by the slightly increase in PSE, see Figure 4.7B), since in the periphery visual acuity decreased and the 1 px closing lines sometimes can be misperceived or not be sufficient to completely suppress the surface completion. Note that we recently tested two different types of

closing-line size (1 px and 4 px) and in both cases no effect of ICs alignment was found (Adriano, Rinaldi & Girelli, 2021).

However, and crucially, the amount of rotation of the inducers and the overall spatial position of the items were exactly the same across Experiment 2 and Experiment 3. Hence, if the change in PSE in the Experiment 2 was simply due to mere inducer orientation statistics, a strong effect of item rotation should have been found also in Experiment 3. Yet, we found only a small (non-significant) effect of inducer alignment in Experiment 3 (see also Adriano, Rinaldi, et al., 2021; Kirjakovski & Matsumoto, 2016), which could be compatible with the findings of Chen and colleagues (2018) but not with the idea that orientation statistics drive the increase in PSE per se. These findings further corroborate the idea that the underestimation effect found in Experiment 1 and Experiment 2 is specifically due to the *strong* binding of the single items in the pairs driven by the (modal) ICs connecting-lines, generated by a boundary system insensitive to the contrast polarity of the inducers, rather than to mere inducers orientation statistics.

4.6. General Discussion

Recent studies in non-symbolic numerical cognition have employed visual illusions as a tool to reveal the exact mechanisms underlying visual perception of numerosity (Adriano et al. 2021; Dormal et al., 2018; Franconeri et al., 2009; He et al., 2009; Picon et al., 2019). Specifically, Kanizsa-like illusory contours have been used to precisely manipulate connectedness-level (or grouping strength) of the individual dots (Adriano et al., 2021; Kirjakovski & Matsumoto, 2016) without changing main low-level features in the scene (e.g., density, occupancy, total area, convex hull, etc). Overall, this line of research indicates that a rapid visual scene segmentation mechanism might be at the root of visual numerosity extraction (e.g., Burr & Ross, 2008; Dehaene & Changeux, 1993; Grossberg & Repin, 2003; Piazza et al., 2004; Stoianov & Zorzi, 2012; Verguts & Fias, 2004), rather than mere continuous cues and/or texture statistics (e.g., Dakin et al., 2011; Gebuis et al., 2016; Leibovich et al., 2017). However, such illusory figures may have involuntary introduced further uncontrolled visual continuous confounds such as the perceived brightness in visual scene. Hence, the observed underestimation reported by previous studies might be potentially explained by uncontrolled continuous cues, and/or as a reduced light sampling of visual input, rather than by the grouping itself of the single inducers into discrete segmented objects.

To shed light on this issue, here we manipulated the brightness level of the illusory contours. Our results clearly indicate that underestimation triggered by ICs lines did not depend on the perceived change in brightness of the illusory surface per se. Rather, the fact that the underestimation is preserved when inducers had opposite contrast polarity, regardless of the color of the inducers (all black or all white) as in Experiment 1, or one black and one white as in Experiment 2, strongly suggests that a mechanism insensitive to contrast direction may drive the numerical underestimation. Finally, in Experiment 3 we found that the underestimation effect was specifically due to the illusory boundary formation and not to other image statistics (e.g., item orientation).

Beside this, we believe that the different change in PSE in Experiment 1 and Experiment 2 (as compared to Experiment 3) suggests that our findings cannot be explained by long-term memory effects. If participants memorized the overall spatial pattern or even the single position of each item in test or reference stimulus, no difference in perceived numerosity (PSE) should be found across ICs conditions in Experiment 1 and Experiment 2, since stimuli pairs had all the same patterns across ICs conditions. Finally, we totally exclude an effect of memory because closed inducers and open inducers (Experiment 2 vs Experiment 3), were also cloned across experiments. In Experiment 3 we

presented the same trials and stimuli pairs (yet, in random order) with exactly the same visual patterns and item position of Experiment 2 (cf. Figure 4 vs Figure 6). Because the pattern of PSE was very different across experiments but, overall, stimuli spatial patterns were equals, we are confident that long-term memory effects cannot account for the overall patterns of results within or between experiments

According to the FACADE model, ICs are encoded by two interacting but complementary streams in early visual cortex (e.g., Grossberg, 2014). In particular, as supported by psychophysical and computational data, ICs are generated by a boundary mechanism *insensitive* to opposite contrast polarity and by a filling-in mechanism *sensitive* to the direction of contrast (Dresp et al., 1996; Grossberg, 2014; Matthews & Welch, 1997). This model might explain why when the notches of the two single inducers are collinear (independently of their polarity), they instantiate the connecting illusory line so that the pac-man shapes are perceived as forming one dumbbell-like object, as the grouping by element connectedness effect would suggest (e.g., Palmer & Rock, 1994). That is, a neural mechanism in visual cortex should trigger the IC-connecting line taking as input the two single separated inducers, thus forcing the two inducers to be processed as a unified single connected object. Neural models suggest that this operation is carried out in neurons with properties similar to a logical AND gate, called bipole grouping cells. Indeed, neurons with similar features, have been found in monkey V2 cortex (Baumann, Van der Zwan, & Peterhans, 1997; von der Heydt, Peterhans, & Baumgartner 1984) and their properties have been confirmed also by later psychophysical work (Field, Hayes, & Hess, 1993; Shipley & Kellman, 1992). Bipole cells can complete boundaries in response to colinear inducers with the same relative contrast and between inducers with opposite relative contrasts with respect to the background, receiving their inputs from complex cells in layer 2/3 of cortical area V1. Complex cells, in turn, pool inputs from simple cells in layer 4 of V1 that have the same preferences for position and orientation, but opposite contrast polarities (e.g., Dresp & Grossberg, 2016). Later neural models and psychophysical data (Kogo, Strecha, Van Gool, & Wagemans, 2010) have highlighted additional specific features and constraints allowing the emergence of the ICs (see also Spehar, 2000).

However, we pinpoint that in the current work the “context” in which we use this classic Kanizsa-like illusion revealed a systematic *underestimation* of numerosity, since the grouped inducers were perceived as belonging to a “whole” single object. Indeed, the illusion used in the current study is a combination of two illusions tapping onto different but strictly related aspects of perceptual organization that have been mostly investigated separately: the grouping principle of

element connectedness defining the entry-level unit of visual perception (e.g., Palmer & Rock, 1994) and the specific rules governing the emergence of the classic Kanizsa illusion and the filling-in process (e.g., Grossberg, 2014). While many psychophysical studies have investigated the rules governing the filling-in and boundary completion in ICs manipulating inducers features such as contrast and shape (Dresp et al., 1996; Grossberg, 2014; Kogo et al., 2010; Matthews & Welch, 1997; Spehar, 2000), little attention has been directed to the unifying effect of this emergent illusory shape over the perceptual organization of the overall figure, formed in this case by the two inducers *and* the connecting-illusory line (e.g., hence forming a dumbbell-like object). Therefore, in our work the IC generated by the two inducers was totally task irrelevant since subjects were instructed to estimate the number of inducer shapes. This striking combination of illusions (e.g., element connectedness driven by ICs) with the specific task-context used, in which inducers were the to-be-counted items, may reveal a form of grouping by element connectedness occurring even though two physically separated surfaces are illusorily connected. Recently, Roelfsema and Houtkamp (2011) suggested a neural model suited to explain the effect of physical connectedness that can also accommodate the case of illusory connectedness and other grouping rules. In particular, when inducer elements are not directly connected by a physical line but by other grouping rules such as good continuation, the model assumes a spread of enhanced neural activity through horizontal connections between neurons tuned to well-aligned contour elements (Field et al., 1993; Grossberg & Raizada, 2000), corresponding to the spread of object-based attention at the psychological level.

In sum, the use of the ICs illusion in the context of a numerical task, in which inducers were task relevant items, suggests that the illusory boundary triggers an overall organization of the inducers into a global shape. This “hidden” grouping effect might have been overlooked in classic research about ICs since inducer features were often manipulated to study the factors underlying the grouping of the aligned inducer-edges triggering the emergence of (task-relevant) illusory shape and brightness. Thus, the task-context itself was not favorable to capture this (context-dependent) grouping mechanism generated by the ICs over the hierarchical organization of the numerosity input units. Indeed, as we emphasize, grouping is not working only at the level of the inducers collinear edges generating the ICs line, but once this line is triggered, the inducers are likely grouped in a coherent whole configurational object (as if the line were a complete physical line) and this whole object is then selected as a single input unit for numerosity computation. Our results are also in line with previous studies in which connectedness manipulation was used. For instance, recent studies

also found an effect of physical connectedness in the early visual cortex, suggesting that numerosity segmentation might start from this stage of visual processing (Fornaciai & Park, 2018, 2021).

Hence, the convergence of two separated lines of research investigating the effects of grouping by element-connectedness and the ICs formation rules, applied in the context of a numerical task, suggests that numerosity perception could be affected by the hierarchical organization of the raw visual input. Similar types of global biases have been also reported for other hierarchically organized objects, such as (global) letters composed of other smaller (local) letters (Navon, 1977). As a consequence, numerosity perception is a promising field to investigate also the effects of Gestalt grouping cues in visual perception (Roelfsema & Houtkamp, 2011). In addition, the effect of connectedness is in line with recent findings suggesting that object-based attention, as well as location-based attention, may modulate the representation of numerosity (Pomè, Thompson, Burr, & Halberda, 2020). Therefore, we pinpoint that our study was carried out to specifically test whether prior works using ICs as connecting lines (Adriano, Rinaldi, et al., 2021; Kirjakovski & Matsumoto, 2016), which used medium numerosities (e.g., 9 – 15) outside the so-called subitizing range (e.g., less than 5), could be explained by the illusory brightness enhancement confounds in the stimuli. It is worth noting that while some studies using estimation tasks argue against a substantial impact of connectedness in the subitizing range at both behavioral and neural level (Porter, Mazza, Garofalo, & Caramazza, 2016; Wurm, Porter, & Caramazza, 2019), others have documented an influence of connectedness over subitizing mechanisms using comparison tasks (He et al., 2015). This suggests that these perceptual manipulations (e.g., connectedness) may depend on several contextual factors such as the experimental task and/or the numerical range used. Indeed, for very large numerosities or dense arrays, the effect of connectedness is weaker or reversed even with comparison tasks (e.g., Anobile, Cicchini, Pomè, & Burr, 2017; Kirjakovski & Matsumoto, 2016). These apparently diverging findings may be explained by the fact that subitizing and ANS are subserved by different cognitive mechanisms (for a review see Hyde, 2011). Visual illusions are thus critical tools to study how visual scene is segmented and to understand the role of low-level features in numerosity processing.

4.7. Conclusions

The current study shows that numerosity perception is not based on continuous cues processing. Our data indeed indicate that increasing illusory connecting lines produces a systematically larger numerosity underestimation effect. This effect is not simply due to physical cues neither to brightness confounds (e.g., filling-in process), but rather explained by a unifying process acting independently of contrast polarity, such as the boundary completion process. In sum, our findings suggest that numerosity is computed based on the rapid segmentation of the visual input into coherent segmented objects, reinforcing the idea that numerosity perception might be biased toward a global organization of the visual input.

Chapter 5

5.1. Discrete numerosity is encoded independently from perceived size.

Chapter adapted from: Adriano, A., Girelli, L., & Rinaldi, L. (2021). *Psychonomic Bulletin & Review*, 1–11. https://doi.org/10.3758/s13423-021-01979-w

5.2. Introduction

Disentangling the contribution of discrete information (e.g., the number of segmented entities in the set or numerosity) from continuous visual features confounded with numerosity (e.g., convex hull, density, area, etc.) represents the main theoretical and experimental challenge to probe which visual mechanisms and sensory features are exploited by the ANS to reach an approximate numerical representation (Gebuis et al., 2016; Leibovich et al., 2017). Visual illusions could be the ideal tool to dissociate the subjective perception of (discrete) numerosity from continuous features because they help to reveal the relationship between physical stimulation (e.g., at the retinal level) and the subjective perception of the visual input. Therefore, they can be used to selectively manipulate a visual feature without compromising other physical visual features in the image (e.g., Picon, Dramkin, & Odic, 2019). For instance, the connectedness illusion has been used to manipulate the level of perceived segmentation of the items in a set, keeping constant the low-level features across connectedness levels (Adriano, Rinaldi, & Girelli, 2021; Adriano, Girelli, & Rinaldi, 2021;Franconeri et al., 2009; Kirjakovski & Matsumoto, 2016). In particular, some of these studies employed Kanizsa-like illusory contour lines (e.g., Nieder, 2002) to connect the dots in the set. Results showed that increasing the illusory connected dot-pairs proportionally reduced the perceived numerosity (i.e., as a function of the number of illusory connections). This result is likely to emerge because the visual system processes two connected dots as a single unified perceptual object (e.g, Anobile et al., 2017; Franconeri et al., 2009), as maintained by the grouping principle of element connectedness (Palmer & Rock, 1994). These findings thus suggest that non-symbolic

numerosity would be extracted from discrete segmented (perceptual) objects rather than from raw low-level features of an unsegmented scene.

By contrast, other studies manipulating the perceived size of continuous features by means of size illusions, bring evidence in favor of the indirect account (Dormal, Larigaldie, Lefèvre, Pesenti, & Andres, 2018; Picon et al., 2019). Size illusions are perceptual phenomena in which the physical size of a stimulus is altered by contextual cues. For instance, Picon and colleagues (2019) contingently manipulated numerosity and perceived size, embedding numerical arrays in the classic Ebbinghaus illusion context. Results showed that participants significantly overestimated the number of dots presented in a *perceived* larger convex hull and underestimated the number of dots presented in the *perceived* smaller convex hull. Accordingly, and in line with indirect accounts, they suggested that numerosity would be mainly encoded through continuous physical features (e.g., convex hull/density).

Despite these previous studies employing different visual illusions seem to reach contradictory conclusions, it is worth noting that they used only one type of illusion at time targeting, in turn, different key visual information in the stimuli. That is, studying the effect of visual illusions *in isolation* does not provide much insight regarding whether i) one type of information (i.e., discrete elements or continuous variables) prevails over the other or ii) both types of information independently contribute to numerosity perception. To this aim, in the present study, we *concurrently* applied two different visual illusions over the same stimuli to shed light on the processing of visual discrete numerosity information and continuous physical features. A similar approach, combining visual illusions in the same stimulus, has been already employed to investigate the extent to which different simultaneous visual distortions may interact affecting the final percept, a condition not uncommon in real world perception and in visual arts such as drawings (e.g., Ni, 1934; Coren & Ward, 1979). In particular, here we employed the Kanizsa illusion to manipulate the perceived item segmentation as well as the Ponzo illusion, a geometrical optical illusion, to manipulate the perceived convex hull/density of the set. We independently modulated the direction of each illusion bias (e.g., underestimation or overestimation) but, crucially, keeping constant at the same time all the physical and contextual cues across key experimental conditions. Hence, illusions were presented in isolation or in a merged condition (e.g., combining the effects of the two illusions).

5.3. Experiment 1: Congruent Illusions

In the Experiment 1, participants performed a number comparison task in which we manipulated the effect of the two illusions obtaining 4 different experimental conditions: one condition without illusions (e.g., baseline), one condition with only the Kanizsa illusion, one condition with only the Ponzo illusion, and one combined (or merged) condition with illusions triggering a bias in the same direction (e.g., both acting toward an underestimation bias). If numerosity is processed independently from continuous magnitudes, we should find the larger underestimation in the combined condition compared to the single illusion conditions. On the contrary, according to the indirect account, the bias in the combined condition should not differ from the bias in the condition with only the Ponzo illusion, as the perceived convex hull/density should play the leading role in driving numerosity estimation.

5.3.1. Materials and methods

5.3.1.1. Participants

Due to Covid-19 restrictions in Italy, the participants were recruited through Pavlovia (www.pavlovia.org), a repository and launch platform allowing online experiments. A total sample of 67 participants (*M*-age = 33.8, *SD* = 11.4, 46 females, 57 right-handed) took part in the study. All participants had normal or correct-to-normal vision and were naïve about the purpose of the experiment. The study was approved by the Local Ethical Committee (protocol N° RM-2020-230).

5.3.1.2. Stimuli and Design

The experimental stimuli were generated off-line by a custom Python/Psychopy script (Peirce, 2007) and were constructed with the same specifications as in Adriano et al. (2021), adding the specific context lines forming the Ponzo illusion.

The whole experimental set was composed of 168 test patterns (42 random spatial patterns cloned across 4 illusion conditions) and of 168 reference patterns (42 random spatial patterns repeated 4 times to match the 4 illusion conditions). The reference patterns always contained the same numerosity (*N*=12), consisting of 12 black “pac-man” like items (diameter = 20 pixels; notch width = 4 pixels; notch length = 10 pixels, measured from the center; RGB = -1, -1, -1) spatially

scattered and randomly rotated at an angle varying across 360° to avoid collinearities and pop-out of illusory contours (ICs). The test patterns contained a variable numerosity, that is, from 9 to 15 "pac-man" like items. Half of the test patterns (*N*=84) were composed by "pac-man" items that were not eliciting any ICs, while in the other half of test patterns (*N*=84) "pac-man" items were purposely aligned to prompt ICs (i.e., the Kanizsa illusion).

Overall, 4 different experimental conditions were designed (see Figure 5.1 for a graphical depiction), according to the specific test pattern employed: a) a no-illusion condition, in which neither the Kanizsa nor the Ponzo illusions were presented (i.e., the "pac-man" items did not trigger any ICs and the sets were embedded in two parallel lines); b) a Kanizsa illusion condition, in which only the effect of items connectedness was manipulated (i.e., the "pac-man" items were aligned to trigger ICs and the sets were embedded in two parallel lines): in this case, an underestimation is thus expected; c) a Ponzo illusion condition, in which only the perceived convex-hull/density of the sets was manipulated (i.e., the "pac-man" items did not trigger any ICs but the sets were embedded in two tilted lines): also here an underestimation is expected, as the test set was always anchored in the larger part of the Ponzo illusory context; d) a combined or merged Ponzo-Kanizsa illusion condition, in which both the effects of items connectedness and the perceived convex-hull/density of the sets were manipulated (i.e., the "pac-man" items were aligned to trigger ICs and the sets were embedded in two tilted lines): in this combined condition, a greater underestimation is expected.

In particular, a first set of 42 test patterns was generated for the no-illusion condition (6 random visual patterns were generated for each of the 7 numerosity values in test stimuli), which were coupled to the 42 reference patterns. In each test pattern of the no-illusion condition set, all the inducers were not aligned (did not trigger ICs). Each of the 42 stimuli pairs of the no-illusion condition (i.e., composed of reference and stimulus patterns) were embedded inside two parallel black lines (width = 2 pixels; RGB = -1, -1, -1) forming a rectangle whose base was 580 pixels and the height 250 pixels placed at the screen center. These two parallel lines did not elicit any illusion (and were used as a control for the Ponzo illusion).

In the Kanizsa illusion condition, to keep constant spatial profiles of test sets from the baseline (i.e., and thus to control continuous variables), each different test pattern for each numerosity of the no-illusion condition was cloned. Thus, we kept constant the spatial position of all the single items in a given test pattern from the no-illusion set. Critically, in this case a sub-set of "pac-man" items was appropriately rotated and aligned to prompt 4 ICs for the Kanizsa condition. The distance between the "pac-man" items that could prompt the required number of ICs for the

connectedness (or Kanizsa) condition was randomly chosen among four possible values (center-to-center distance = 22, 25, 28, and 31 pixels). In this way, the 42 different reference patterns were associated with the same spatial pattern of test stimuli across the no-illusion and the Kanizsa illusion condition. In both conditions test and reference stimuli were embedded inside two parallel black lines, so that no Ponzo illusion was prompted.

Then, these two conditions were cloned and drawn embedded in the Ponzo illusion context, thus generating stimuli pairs for the Ponzo illusion condition and the combined Ponzo-Kanizsa illusion condition. The Ponzo illusion was elicited by two tilted black lines (width = 2 pixels; RGB = -1, -1, -1) forming the legs of an isosceles trapezoid whose virtual longer base was 300 pixels length and whose shorter base was 250 pixels length (distance between the bases of 580 pixels). Note that in the experiment, the relative positions of the reference and the test stimuli were randomized between the left and right side. Yet, the test set was always anchored in the larger part of the Ponzo illusion context, which was randomized in accordance with the position of the test stimulus, so that when the test stimulus appeared to the right, the Ponzo illusion context was drawn with the larger side on the right side.

All the patterns in the four experimental conditions were drawn on a grey background (RGB = 0, 0, 0) and reference and test stimuli were projected within two virtual squared panels (240 x 240 pixels) centered at ±156 pixels from the screen center. Furthermore, we constrained the single “pac-man” items in test and in reference stimuli to be distant at least 20 pixels from the 4 virtual square edges and to not overlap with each other (minimum center-to-center distance = 22 pixels).

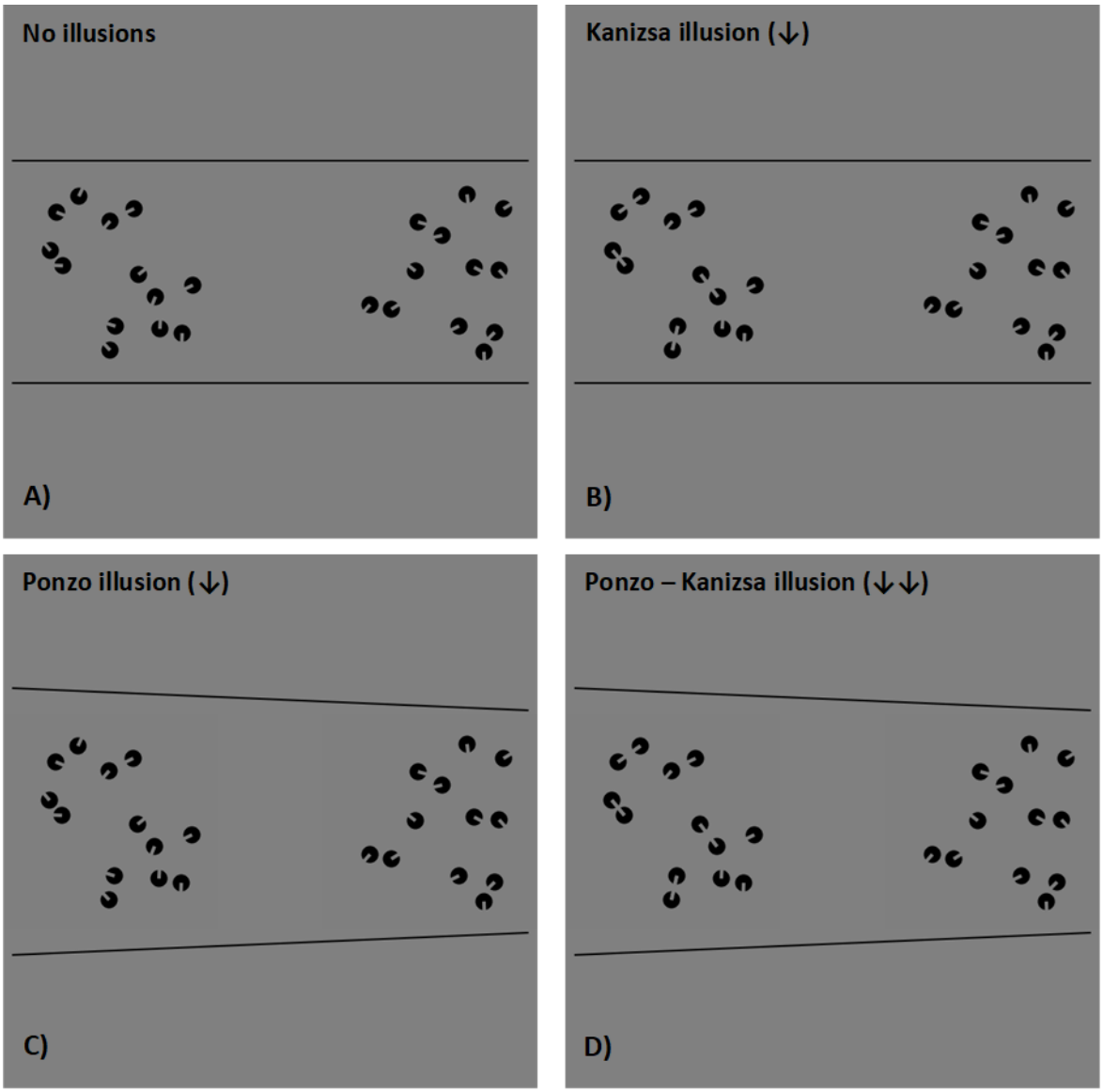


Figure 5.1: The four experimental conditions of Experiment 1: a) the no-illusion condition, in which neither the Kanizsa nor the Ponzo illusions were presented; b) the Kanizsa illusion condition, in which only the effect of items connectedness was manipulated; c) the Ponzo illusion condition, in which the perceived convex-hull/density of the sets was manipulated; d) the Ponzo-Kanizsa merged illusion, combining the effects of these last two conditions. In the example reported, the test set (9 to 15 items) is positioned on the left side of the screen (i.e., and hence in the larger side of the Ponzo illusion), while the reference (always 12 items) on the right side (i.e., and hence in the smaller side of the Ponzo illusion). The small arrows represent the direction of the predicted bias of each illusion: in particular, an underestimation is expected for both the Kanizsa and the Ponzo illusions; moreover, if the effects of the two illusions (and thus the effects of segmentation mechanisms and continuous variables) would be additive, the greater underestimation should be observed in the merged condition.

5.3.1.3. Procedure

The stimuli were presented by means of an on-line Psychopy routine (Peirce, 2007) and all the experimental materials (stimuli, etc.) were downloaded and stored on the computer of each participant. The general procedure was explained to each participant before starting the experiment by means of detailed instructions provided on the display. No information about the illusions was given to the participant.

The participants performed a two-alternative forced choice task, in which they were asked to choose the set containing more objects between two rapidly presented visual patterns by pressing the corresponding keys on the keyboard. The experimental phase was preceded by a brief training composed of 24 trials (6 trials for each of the 4 illusion contexts) to allow the subject to familiarize with the task. In the training phase, we presented only the reference patterns vs. the test pattern with 9 items. Each experimental trial started with a middle-grey background (RGB = 0, 0, 0) lasting 1000 ms, followed by a black fixation cross (Font: Times; Size: 16 pixels; RGB = -1, -1, -1) projected for 1000 ms, and then, two collections of dots appeared at the left and right of the center of the screen (i.e., the two collections were centered at ±156 pixels from the screen center) for additional 400 ms (Figure 5.2). The side of the reference and test patterns was counterbalanced and randomized across trials. Test set was anchored to the larger side of the Ponzo illusion, which was randomized accordingly to the left or to the right, following the test stimulus side. After the stimuli offset, an empty screen (RGB = 0, 0, 0) was presented until the participant's answer. The subjects could select the stimulus by pressing the appropriate key with their left or right index finger ("F" key for the left stimulus and "J" key for the right stimulus).

Response time was not restricted, but we emphasized in the instructions to answer as fast as possible. After the practice session, two counterbalanced blocks (i.e., across participants) composed of 168 randomly ordered trials were presented, for a total of 336 experimental trials (12 trials for each of the 7 numerosities across the 4 illusion contexts), separated by a self-paced pause at the half of the whole session. The whole experiment lasted around 15-20 min.

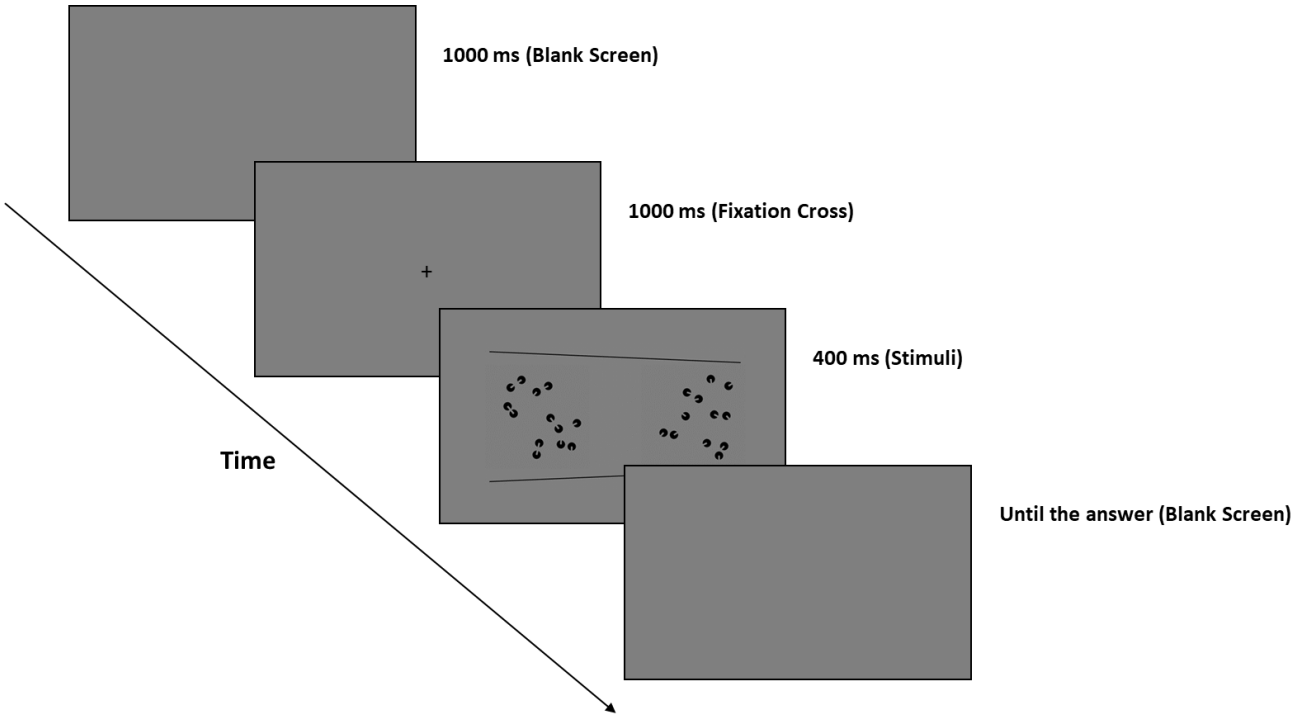


Figure 5.2. The numerical comparison task. Participants had to indicate the numerically larger between the two collections of dots by pressing the corresponding left or right key.

5.3.1.4. Data analysis

The data were analyzed with *R-Studio* (2018, v. 3.6.2; http://www.rstudio.com/) and *Jamovi* (2019, v. 1.1.5; https://www.jamovi.org) softwares. Psychometric functions for each condition were generated by fitting Gaussian cumulative distribution functions to the data and parameters were estimated with a parametric approach based on maximum likelihood method, using *Quickpsy* package for *R* (Linares & López-Moliner, 2016). In order to minimize biases in estimating the psychometric function parameters, we fitted the psychometric curves taking into account the typical lapse in performance (e.g., missing a trial, finger-errors) by allowing the value of the guess rate (γ) and lapse rate (λ) parameters to vary in the default range of 0 – 0.05 (Wichmann & Hill, 2001).

To investigate the effect of the illusions over perceived numerosity, we calculated the point of subjective equality (PSE) for each illusion condition as a function of the numerosity in test set: that is, the number of dots in test patterns required in order to be subjectively judjed as equal to

the the reference patterns (12 items). The 50% of the chosen test patterns was set as threshold level. The 95% confidence intervals of individual PSEs were estimated running 200 bootstrap resampling of the data. Furthermore, as an index of the precision of the numerical discrimination and to confirm that the performance follows the Weber's law (e.g., JND/N = *k*) we calculated the Coeffienct of Variation (CoV; Whalen et al., 1999), as the ratio between the standard deviation (SD) and the PSE of the psichometric functions for each illusion condition. Reaction times (RTs) for each illusions condition were also recorded. RTs data were logarithmically transformed and responses whose latencies fell outside of 1.5 times the interquartile range of the distribution were discarded (a total 4.89 % of the trials were discarded from RTs data). Two separated one-way repeated measures ANOVAs were performed with the experimental condition (No Illusion, Kanizsa, Ponzo, Ponzo-Kanizsa) as within-subjects factor and with the mean PSE or the mean CoV as dependent variables. Furthermore, we performed a 4 x 4 repeated measures ANOVA with the absolute numerical distance between reference and test stimuli (0, 1, 2, 3) and the experimental condition (No Illusion, Kanizsa, Ponzo, Ponzo-Kanizsa) as within-subjects factors and the mean RTs as dependent variable. The Greenhouse-Geisser epsilon (ε) correction for violation of sphericity was applied when needed and original *F*, *df* and corrected *p*-values were reported. Frequentist analyses were also accompanied by respective Bayesian analysis in the case of non-significant results.

5.3.2. Results and Discussion of Experiment 1

The analysis on the PSE (i.e., the higher the PSE the greater the underestimation bias) showed a significant effect of the experimental condition, $F(3, 198) = 55.1$, $\varepsilon = .89$, $p < .001$, $\eta^2_p = .45$, Figure 5.3A and Figure 5.3B. Post-hoc comparisons (Bonferroni-Holm correction) revealed a significant difference between the baseline (mean PSE ± SD, 12.025 ± 0.49) and the Kanizsa condition (12.24 ± 0.39), $t(198) = -2.79$, $p = .012$, the baseline and the Ponzo condition (12.74 ± 0.41), $t(198) = -8.99$, $p < .001$, the baseline and the combined condition (12.92 ± 0.62), $t(198) = -11.26$, $p < .001$, as well as between the Kanizsa condition and the Ponzo condition, $t(198) = -6.2$, $p < .001$, between the Kanizsa condition and the combined condition, $t(198) = -8.47$, $p < .001$, and crucially, between the Ponzo condition and the combined Ponzo-Kanizsa condition, $t(198) = -2.27$, $p = .024$. No significant effect of the experimental illusion condition was found over the mean CoV, $F(3, 198)$

= 1.06, p = .36, η^2_p = .016, Figure 5.3C. To further quantify the magnitude of this null effect we also computed the Bayes Factor (BF), and we found strong evidence in favor of the null hypothesis, BF_{10} = .066. We also evaluated the relationship between individual CoVs across the illusion conditions. Results showed a significant correlation with a strong positive relationship for the acuity for each pairwise comparison (all r Pearson's coefficients between .65 – .76, all p < .001, Figure S5.1 and Table S5.1), suggesting that a common sensory mechanism may drive the discrimination performance across the illusion conditions.

Finally, the ANOVA on RTs showed no significant main effect of the illusion condition on RTs, F(3, 198) = 0.433, ε = .88, p = .70, η^2_p = .007; BF_{10} = .004. Figure 5.3D and Table S5.2. A significant main effect of the absolute numerical distance between test and reference was found, F(3, 198) = 43.73, ε = .87, p < .001, η^2_p = .87. Trend analysis showed a significant linear decrease of RTs when numerical distance increases, t(198) = -10.11, p < .001. Lastly, no significant interaction between variables was found, F(9, 594) = 0.56, ε = .73, p = .778, η^2_p = .008. Results from Bayesian analysis suggest strong evidence against including the interaction, BF_{10} = .0008 (e.g., full model with interaction compared to the model with only main effects, Table S5.2).

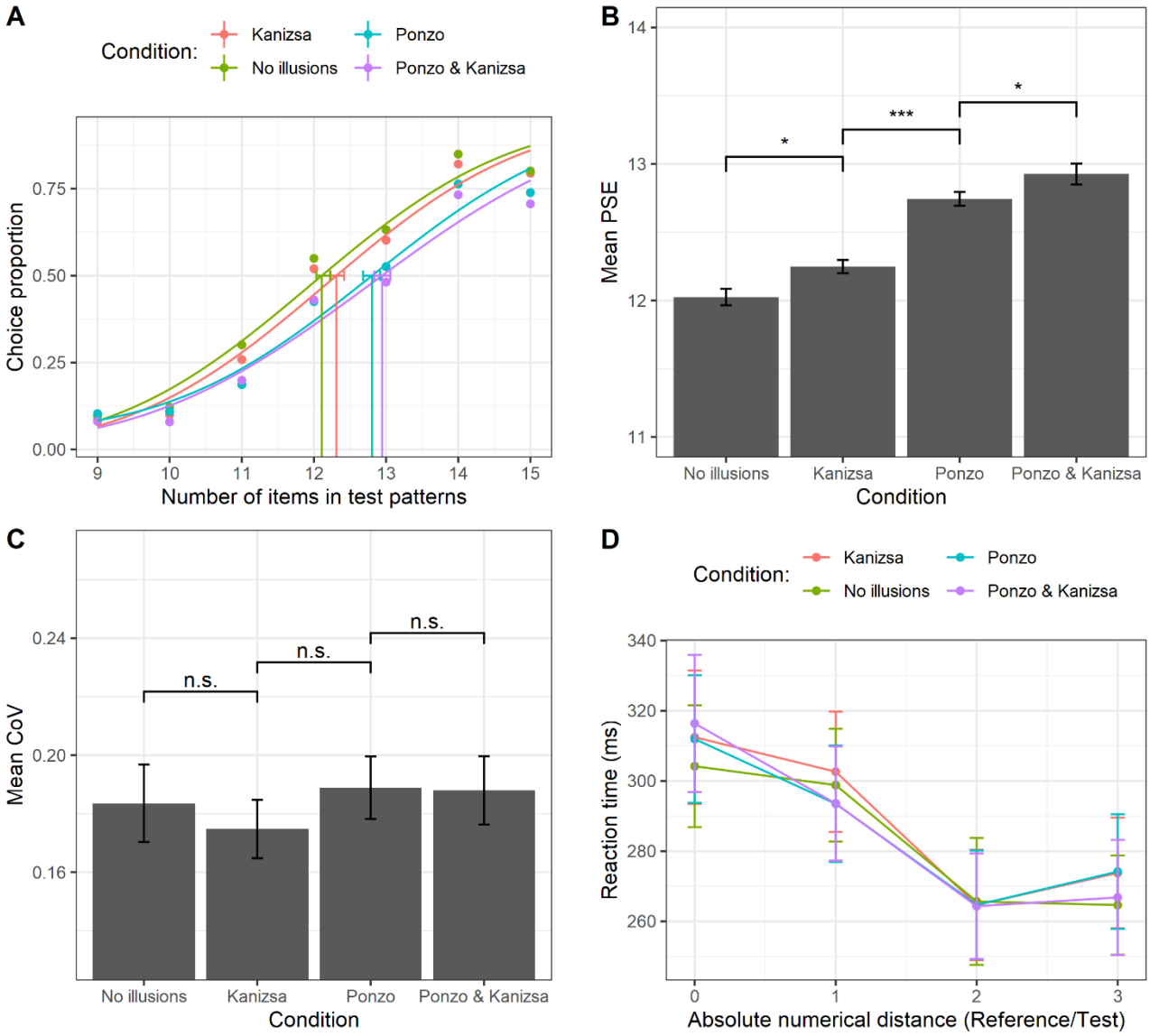


Figure 5.3. (A) Psychometric functions obtained fitting Gaussian cumulative distribution function (for each experimental condition) pooling over the aggregate data of all the subjects. Please note that this graph is reported to illustrate the statistical technic, but all subsequent analyses were done with similar functions over individual subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) Mean PSE as a function of the experimental condition. (C) Mean CoV as a function of the experimental condition. (D) Mean RTs as a function of the experimental condition and the absolute numerical distance between reference and test stimuli. The error bars represent ± 1 standard error of the mean (SEM). * $p < .05$; ** $p < .01$; *** $p < .001$; ns = non-significant.

5.4. Experiment 2: Incongruent Illusions

In the Experiment 2, participants performed the same number comparison task of Experiment 1 but in the combined condition the biases of the two illusions were modulated in a conflicting direction. That is, the Ponzo illusion triggered an overestimation bias, while the Kanizsa illusion an underestimation bias. If discrete information is processed independently from continuous features, when the two individual illusions are combined over the same stimulus in a conflicting direction, we should expect participants' bias to be halfway as compared to the single illusory conditions, indicating that the two illusions compete against each other.

5.4.1. Materials and methods

5.4.1.1. Participants

A new sample of 68 participants (*M*-age = 30.9, *SD* = 12.9, 49 females, 55 right-handed) was recruited for this study, which was also performed on-line due to Covid-19 restrictions.

5.4.1.2. Stimuli and Procedure

The stimuli generation and the procedure were identical to the Experiment 1. The only difference was that test stimuli were anchored to the smaller side of the Ponzo context (Figure 4). The relative side of the reference and of the test patterns was counterbalanced and randomized across trials. The smaller side of the Ponzo illusion, was randomly presented to the left or to the right, according to the test stimulus side. In this case, therefore, the directions of the biases triggered by the Ponzo and the Kanizsa illusion were opposite: while the Ponzo illusion should trigger an overestimation as compared to the reference, the Kanizsa illusion should trigger an underestimation bias.

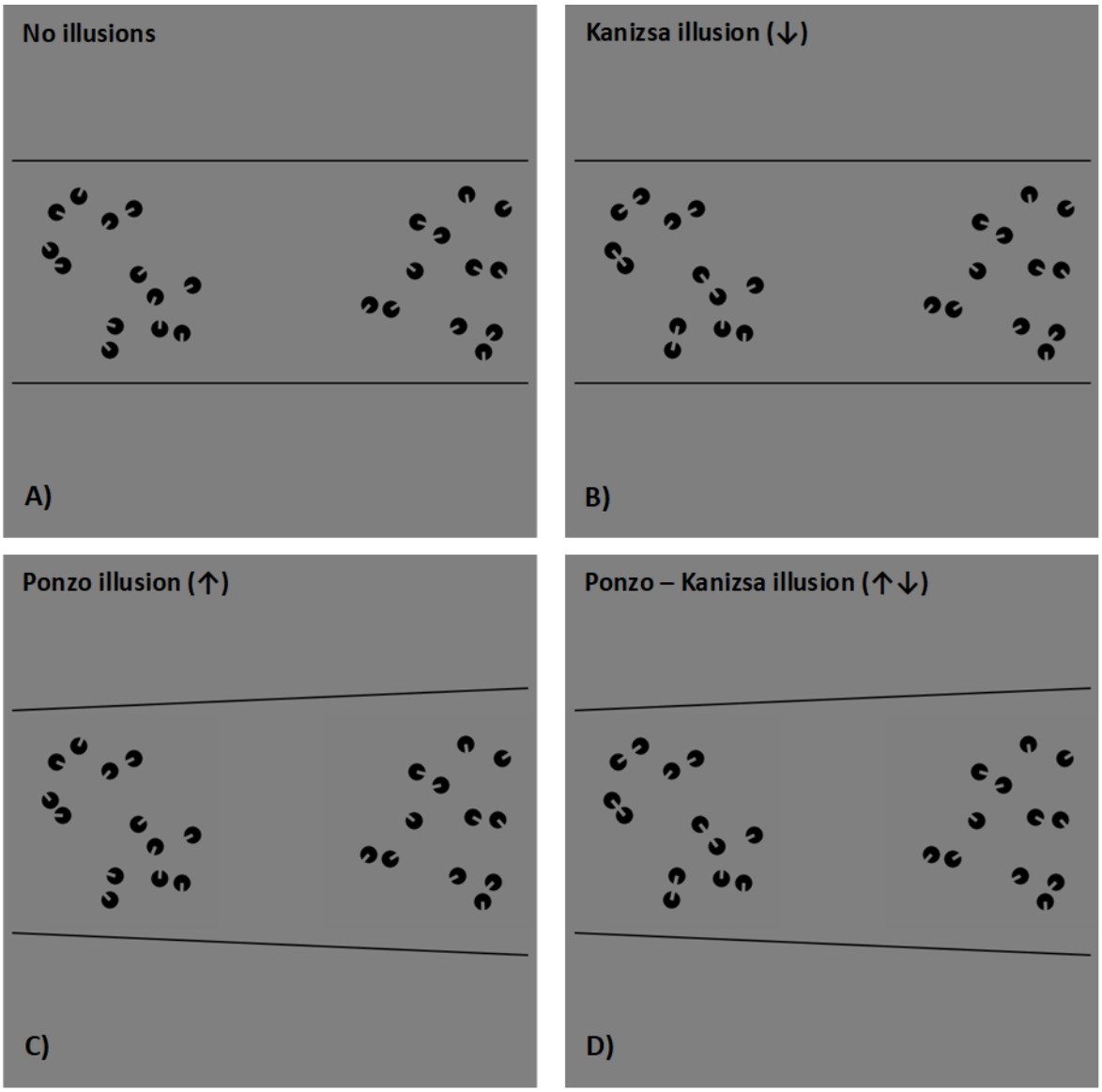


Figure 5.4. The four experimental conditions of Experiment 2: a) the no-illusion condition, in which neither the Kanizsa nor the Ponzo illusions were presented; b) the Kanizsa illusion condition, in which only the effect of items connectedness was manipulated; c) the Ponzo illusion condition, in which the perceived convex-hull/density of the sets was manipulated; d) the Ponzo-Kanizsa merged illusion, combining the effects of these last two conditions. The only difference with Experiment 1 was that the direction of the biases triggered by the Ponzo and the Kanizsa illusion were opposite, with the former that should elicit an overestimation and the latter an underestimation.

5.4.2. Results and Discussion of Experiment 2

Data were analyzed analogously to Experiment 1. The ANOVA on PSE showed a significant effect of the experimental condition, $F(3, 201) = 36.3$, $p < .001$, $\eta^2_p = .35$, Figure 5.5A and Figure 5.5B. Post-hoc comparisons (Bonferroni-Holm correction) revealed a significant difference between the baseline (12.22 ± 0.58) and the Kanizsa condition (12.47 ± 0.56), $t(201) = -2.69$, $p = .015$, the baseline and the Ponzo condition (11.58 ± 0.61), $t(201) = 6.74$, $p < .001$, the baseline and the combined Ponzo- Kanizsa condition, (11.79 ± 0.55), $t(201) = -4.47$, $p < .001$, between the Kanizsa condition and the Ponzo condition, $t(201) = 9.43$, $p < .001$, the Kanizsa condition and the combined condition, $t(201) = 7.17$, $p < .001$, and, crucially, between the Ponzo condition and the combined Ponzo-Kanizsa, $t(201) = -2.22$, $p = .025$. As expected by direct account of numerosity, PSE of the combined condition was between the PSE of the illusion in isolations.

No significant effect of the experimental condition was found over the mean CoV, $F(3, 201) = 0.83$, $\varepsilon = .85$, $p = .46$, $\eta^2_p = .012$, Figure 5.5C. Bayesian analysis suggests strong evidence in favor of the null hypothesis, $BF_{10} = .049$. We also evaluated the relationship between individual CoVs across the experimental conditions. As in Experiment 1, results showed a significant correlation with a strong positive relationship for the acuity for each pairwise comparison (all *r* Pearson's coefficients between .77 – .90*,* all $p < .001$, Figure S5.2 and Table S5.2).

Furthermore, as in the previous experiment no significant main effect of the experimental condition was found over RTs (7 % of the data were discarded), $F(3, 201) = 0.28$, $\varepsilon = .90$, $p = .81$, $\eta^2_p = .004$; $BF_{10} = .004$, Figure 5.5D and Table S5.4. We found a significant main effect of the absolute distance, $F(3, 201) = 14.14$, $\varepsilon = .86$, $p < .001$, $\eta^2_p = .17$. A significant decreasing linear trend was also found over the absolute numerical distance, $t(201) = -4.84$, $p < .001$. Finally, no significant interaction between variables was found, $F(9, 603) = 0.80$, $\varepsilon = .81$, $p = .58$, $\eta^2_p = .012$. Results from Bayesian analysis suggest very strong evidence against including the interaction, $BF_{10} = .0015$ (e.g., full model with interaction compared to the model with only main effects, Table S5.4).

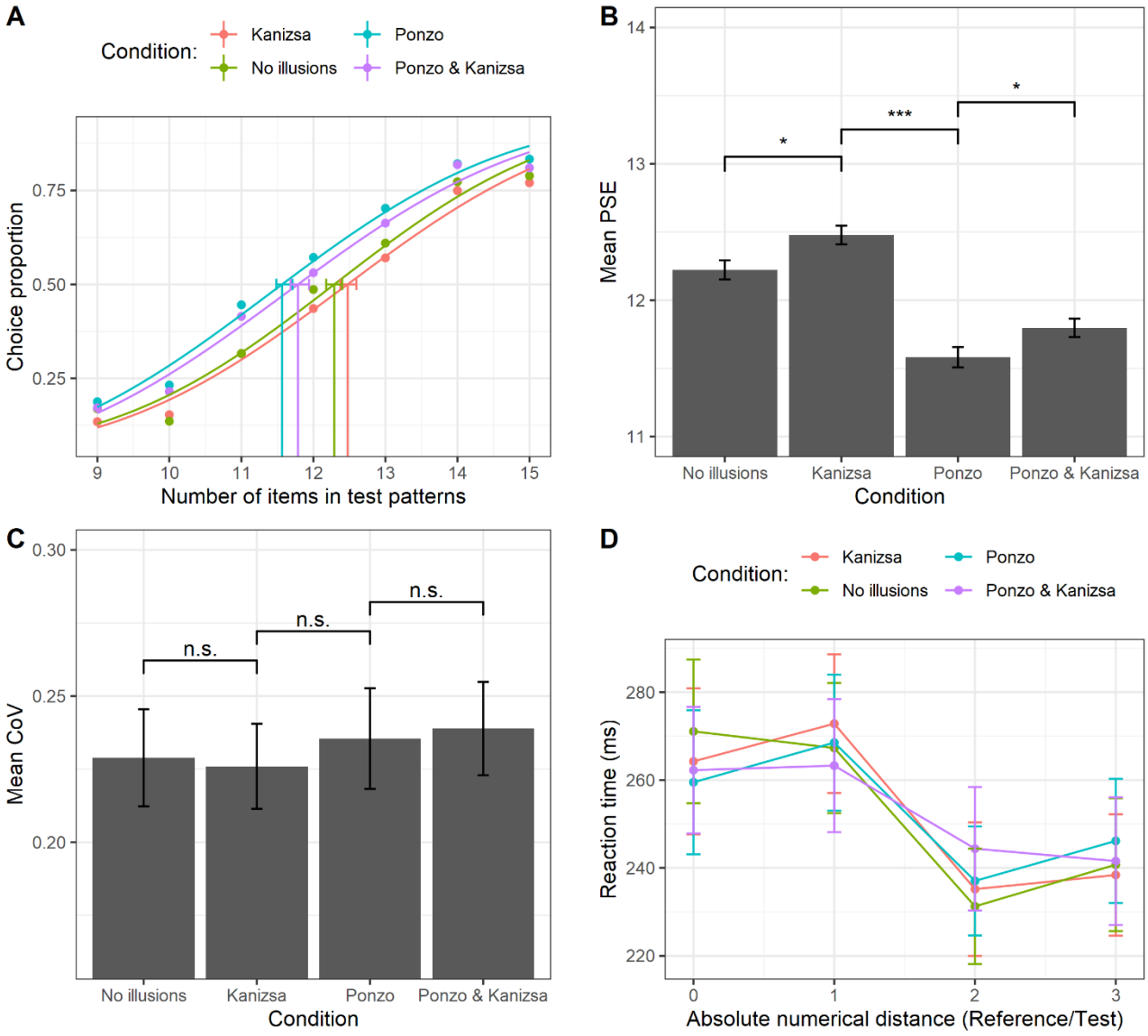


Figure 5.5. (A) Psychometric functions obtained fitting Gaussian cumulative distribution function (for each experimental condition) pooling over the aggregate data of all the subjects. Please note that this graph is reported to illustrate the statistical technic, but all subsequent analyses were done with similar functions over individual subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) Mean PSE as a function of the experimental condition. (C) Mean CoV as a function of the experimental condition. (D) Mean RTs as a function of the experimental condition and the absolute numerical distance between reference and test stimuli. The error bars represent ± 1 standard error of the mean (SEM). * $p < .05$; ** $p < .01$; *** $p < .001$; ns = non-significant.

5.5. General Discussion

In this study, across two different experiments, we document a clear role of visual segmentation mechanisms in discrete numerosity processing (e.g., Burr & Ross, 2008; Franconeri et al., 2009; Piazza et al., 2004; Verguts & Fias, 2004). Indeed, we found that ICs connections in the Kanizsa condition, when manipulated in isolation, led to a numerical underestimation bias in both experiments, as more items needed to be present in the test stimulus in order to be judged as numerically equal to the reference (e.g., increase in the PSEs). These data suggest that the pair of connected objects may form a unity which is selected as input for numerosity (e.g., two dots connected by the illusory contour would be processed as one), hence triggering the change in PSE (underestimation) when the 4 pairs of dots were connected by ICs (see also Adriano, Rinaldi & Girelli, 2021). Notably, this was found despite continuous features were kept constant across baseline and Kanizsa condition, as our test stimuli had the same item spacing, the same total contour (e.g., high spatial frequency) and the same object size and convex hull (e.g., low spatial frequency), thus challenging current alternative views maintaining a key role of these variables (e.g., Allik & Tuulmets, 1991; Dakin et al., 2011; Durgin, 2008; Gebuis et al., 2016).

However, we also found a clear bias triggered by the Ponzo illusion when this illusion was presented in isolation. In line with previous studies using the Ebbinghaus illusion, we found that the Ponzo illusion led to an underestimation (Experiment 1) or an overestimation (Experiment 2) of test numerosity depending on the context in which it was placed (see also Picon et al., 2019). These findings undoubtedly show that also the (perceived) size of the convex hull/density of the set is taken into account for decisions during the comparison task, confirming previous studies (e.g., Picon et al., 2019). Indeed, while these two illusions tap on different perceptual (i.e., the Ponzo illusion induces the perception of three-dimensional depth/distance information, while the Ebbinghaus illusion do not) and neural mechanisms (e.g., Song, Schwarzkopf, & Rees, 2011), the behavioral effects found are very similar in the context of non-symbolic comparison tasks. All these works using size illusions (Dormal et al., 2018; Picon et al., 2019) are also in line with studies showing that visual adaptation to size (e.g., adapter stimuli were discs of different size) affects subsequent perceived numerosity. However, such a size adaptation was stronger only for numerically very large arrays (Zimmermann & Fink, 2016), compatibly with the idea that density and size may be prominent cues only for denser textured stimuli (e.g., Dakin et al., 2011) but not for sparse/lower numerosities (e.g., Anobile et al., 2017). Additionally, Anobile and colleagues (2018) recently asked participants to

perform both a size and a numerosity adaptation task and found that neither discrimination thresholds nor adaptation strength correlate with each other. Crucially, our results showed that when both the Ponzo and the Kanizsa illusions were combined over the same physical stimulus, the joint effect varied according to their bias direction. More precisely, biases of each illusion summated (i.e., largest underestimation as compared to the condition in which only one illusion was presented) in Experiment 1, while they averaged and competed against each other in Experiment 2. These findings can be explained if we assume that when illusions are combined a "discrete" information is still actually processed independently from continuous features. This provides clear evidence against the views maintaining that perceived numerosity is simply the result of weighting a variety of continuous visual properties (Gebuis & Reynvoet 2012b). Rather, our study indicates that both discrete elements and continuous magnitudes can simultaneously affect perceived numerosity and influence the behavior. More critically, the PSE pattern of the *combined* conditions indicates that participants integrate the bias induced by the Kanizsa-connectedness (underestimation) illusion with the bias induced by the Ponzo illusion (either overestimation or underestimation, depending on the experimental manipulation). Indeed, if participants actually ignored discrete numerosity, no difference should be found between the *combined* condition and the Ponzo condition in isolation, since in both conditions we have exactly the same continuous cues (as well as between the no-illusion condition and the Kanizsa condition). This strongly suggests that in the merged condition there is a *combined* effect of both information, namely of discrete (manipulated by ICs) and continuous information (manipulated by the Ponzo illusion). Notably, neither of the two types of information dominated the other, hence no factor was lost or ignored, they were simply combined (according to the direction of biases in each experiment).

Yet, contrary to the holistic view according to which numerosity is processed along with other perceptual variables to construct a sense of magnitude (Leibovich et al., 2017), our data suggest that both discrete elements and continuous variables would be independently integrated together to guide the behavior. Evidence for this integrative process comes from the pattern of results that we observed here over the PSEs across experimental illusion conditions, suggesting that two independent magnitude information might be at play. The fact that, as expected by Weber's law, CoV (e.g., numerical acuity) and RTs were constant across illusion conditions, strongly rules out task difficulty as a possible confounding variable accounting for our findings (e.g., change in PSE across conditions), rather suggesting a genuine equal perceptual discriminability of stimuli across illusion contexts. However, the strong correlations across experimental conditions of individual

precision (indexed by the CoV) and the fact that CoV was stable across illusion conditions also suggest that a common sensory mechanism may operate and drive the discrimination performance in the numerical task, following the Weber's law. The independence between numerosity and size information is also indirectly supported by studies showing that both continuous physical dimensions (e.g., item size or cumulative surface) and discrete number information may be automatically extracted in Stroop-like tasks even when they are irrelevant to the task (e.g., Hurewitz et al., 2006; Nys & Content, 2012) or when dot arrays are passively viewed (Van Rinsveld et al., 2020), and they may interact or compete for behavioral control perhaps in a late decisional stage (Franconeri et al., 2009; Leibovich & Henik, 2014). Furthermore, it has been shown that discrimination threshold for numerosity, but not for size judgement, is impaired in dyscalculic subjects (Anobile, Cicchini, Gasperini, & Burr, 2018), with other studies reporting that arithmetical education selectively improves acuity in non-symbolic numerical discrimination but not in size discrimination (Piazza, Pica, Izard, Spelke, & Dehaene, 2013). These studies more broadly challenge the idea that continuous cues could be at the core of the development of mathematical abilities.

5.6. Conclusions

In conclusion, this study challenges recent theoretical accounts according to which people would not extract numerosity independently from other continuous magnitudes (Gebuis & Reynvoet 2012a, b; Gebuis et al., 2016). Our study, indeed, testifies the existence of a distinct sense of number that allows perceiving discrete numerosities information exploiting segmentation and perceptual organization, but integrating also other features of the visual input, including continuous magnitudes information such as size. That is, we demonstrate that subjective numerosity could be the result of a flexible combination between continuous and discrete information from the visual scene. These findings indirectly support the hypothesis of a general mechanism that allows for processing of both discrete (i.e., number) and continuous dimensions (i.e., space) in parietal areas (e.g., Walsh, 2003), and points to the need of more comprehensive theoretical views that should account for the operations by which both discrete elements and continuous variables signals are computed and integrated together as relevant cues for extracting numerosity information from the visual stream (e.g., Cantrell & Smith, 2013).

Chapter 6

6.1. Non-symbolic numerosity encoding escapes spatial frequency equalization

Chapter adapted from: Adriano, A., Girelli, L., & Rinaldi, L. (2021). *Psychological Research*, 1–14. https://doi.org/10.1007/s00426-020-01458-2

6.2. Introduction

Disentangling numerosity information from continuous visual features confounded in the stimulus (e.g., spatial frequencies, luminance, density, etc) represents a crucial empirical step to understand the visual mechanisms underlying the ANS.

However, so far, only one study carefully controlled for the possible effects of spatial frequency amplitude (and luminance) on visual enumeration making these features completely uninformative about the target numerosity across trials. In particular, Railo and colleagues (Railo, Karhu, Mast, Pesonen, & Koivisto, 2016) reported that estimation time and error rate of human figures presented in a display increase with the target numerosity even when spatial frequency content of the stimuli were equalized across numerosities. Yet, this study used a very limited numerical range (e.g., 1-6 items) and did not control for the possible additional effects of other low-level visual features (e.g., density, convex hull, etc.).

On these grounds, in the present study we systematically investigated whether non-symbolic number processing of a moderate numerosity array (e.g., 9 to 15 items) operates over summary statistics of an unsegmented scene, such as the raw spatial frequency content of stimuli (Dakin et al., 2011; Durgin, 2008) or rather over segmented visual objects (Burr & Ross, 2008a; Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012). We designed two experiments using the main used numerical tasks: the comparison (Experiment 1) and the estimation task (Experiment 2). We manipulated the set numerosity and the connections among some items connecting them with ICs-lines. Crucially, all the stimuli were equalized for spatial frequency (SF) spectrum and luminance, hence these features were completely neutralized as cue for the numerosity in the sets. We

predicted that, if numerosity is derived from the raw spatial frequency content of the image, when spatial frequency content is not informative about the numerosity (e.g., spatial frequencies amplitudes are matched across numerical stimuli) no behavioral signature of the ANS (e.g., numerical distance effect, scalar variability) should be found in the two tasks. In addition, since spatial frequency amplitude profile is kept constant across connectedness levels, numerosity judgments should not vary when the number of illusory connections (or item pairs) is manipulated, in both experiments.

6.3. Experiment 1: Comparison Task

In the Experiment 1 we specifically tested whether *numerical distance effect* (e.g., Buckley & Gillman, 1974) is preserved when spatial frequency content of the stimuli is not predictive about the numerosity. In addition, we tested whether the perceived numerosity is affected by the degree of object segmentation manipulating the connections between items with Kanizsa-like illusory lines (e.g., Adriano, Rinaldi, & Girelli, under review; Kirjakovski & Matsumoto, 2016). Subjects were tested in a non-symbolic comparison task in which we concurrently manipulated the numerical distance between a reference (always 12 items) and a test stimulus (9 to 15 items) and the number of illusory connections (0, 2 or 4 ICs) in the test. All continuous features (e.g., density, convex hull) were kept constant across ICs levels. Crucially, all the stimuli pairs were matched for spatial frequency. According to Dakin's model, we expect that performance in the discrimination task should be independent from the numerical distance and the number of ICs in the stimuli.

6.3.1. Materials and methods

6.3.1.1. Participants

A sample of 18 participants (16 females) took part in the study. The mean age was 22.77 years (*SD* = 4.25), they were all undergraduate or postgraduate students from the University of Milano-Bicocca. Handedness was assessed by asking participants which hand they typically used for writing: a total of 17 subjects out of 18 were classified as right-handed. All participants had normal or correct-to-normal vision and were naïve about the purpose of the experiment. An informed consent document was signed before the experiment began and the study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Local Ethical Committee (protocol N° RM-2020-230).

6.3.1.2. Stimuli

The stimuli were projected on an Acer AL1716 17" LCD monitor (display area: 338 x 270 mm, 1280 x 1024 pixels, 60 Hz) connected to a pc desktop Olidata (AMD Athlon X2 240, 2.80 GHz, 4 GB ram, Windows® 7). The stimuli were randomly generated off-line by a custom Python/Psychopy script and projected by means of a Psychopy routine (Peirce, 2007). Stimuli (Figure 6.1) were constructed with the same specifications as in Kirjakovski and Matsumoto (2016), but with specific

changes in the generation method to accommodate our control of continuous variables (see Experiment 1, *Chapter 3*).

Following a similar methodology of Railo et al. (2016), visual low-level statistical properties (power spectrum and luminance histogram) were fully equalized processing all the experimental stimuli with the SHINE toolbox for MatLab (Willenbockel et al., 2010), which allows to match both the Fourier amplitude spectrum and the luminance histogram across all the input images. The whole set of 224 stimuli (56 references and 56 test stimuli for each ICs condition) originally generated, were submitted to an iterative algorithm (10 repetitions) to jointly match luminance histograms (histMatch function, which matches mean luminance, contrast, skew, etc.) and Fourier amplitude spectra (specMatch function, which matches spatial frequency and orientations). Both Fourier amplitude and histogram matching are performed iteratively to reach a higher degree of matching of both features across output stimuli. As can be observed from Figure 6.2 and Figure S6.1, indeed, stimuli have similar spatial frequencies amplitude spectrum and luminance profile across numerosities and ICs conditions.

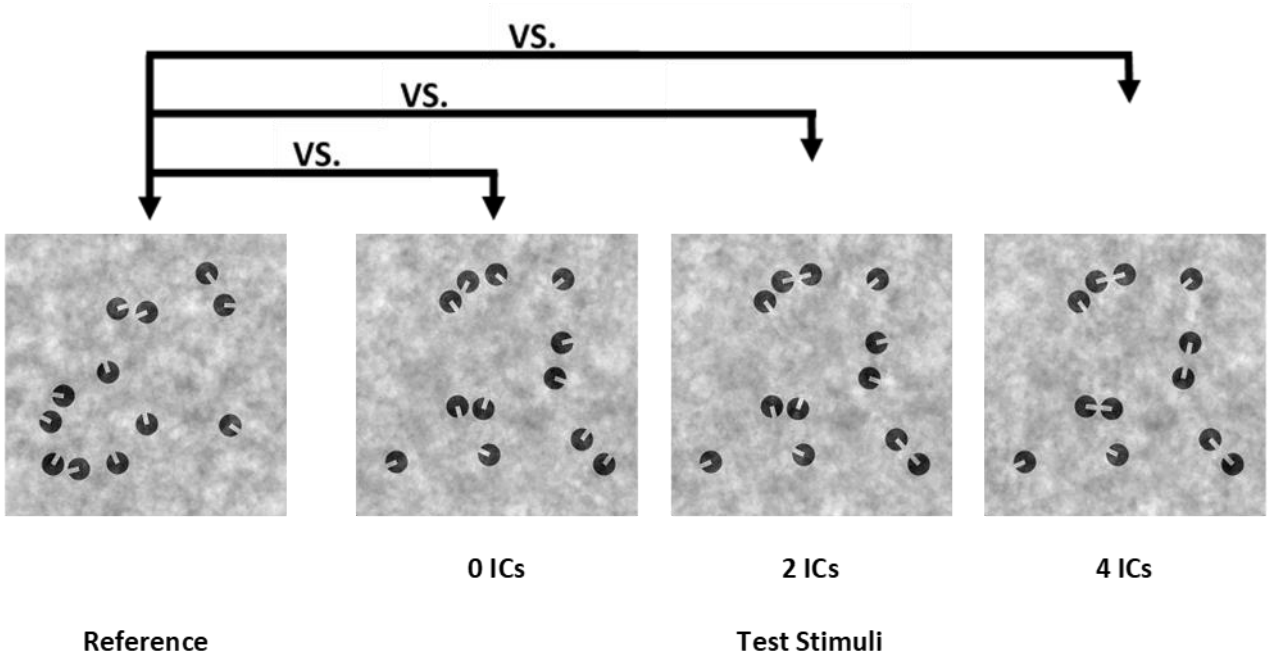


Figure 6.1: The reference stimulus was always composed of 12 items. Test stimuli varied from 9 to 15 items and contained 0 ICs, 2 ICs or 4 ICs. Black arrows indicate that a given reference pattern was associated with the same spatial pattern across the 3 levels of connectedness of test stimuli. All the test stimuli had the same convex hull and density when averaged across the levels of connectedness (as in the examples depicted). All the test and reference stimuli were equalized for spatial frequencies and luminance histograms profile (See also Figure 6.2 and Figure S6.1).

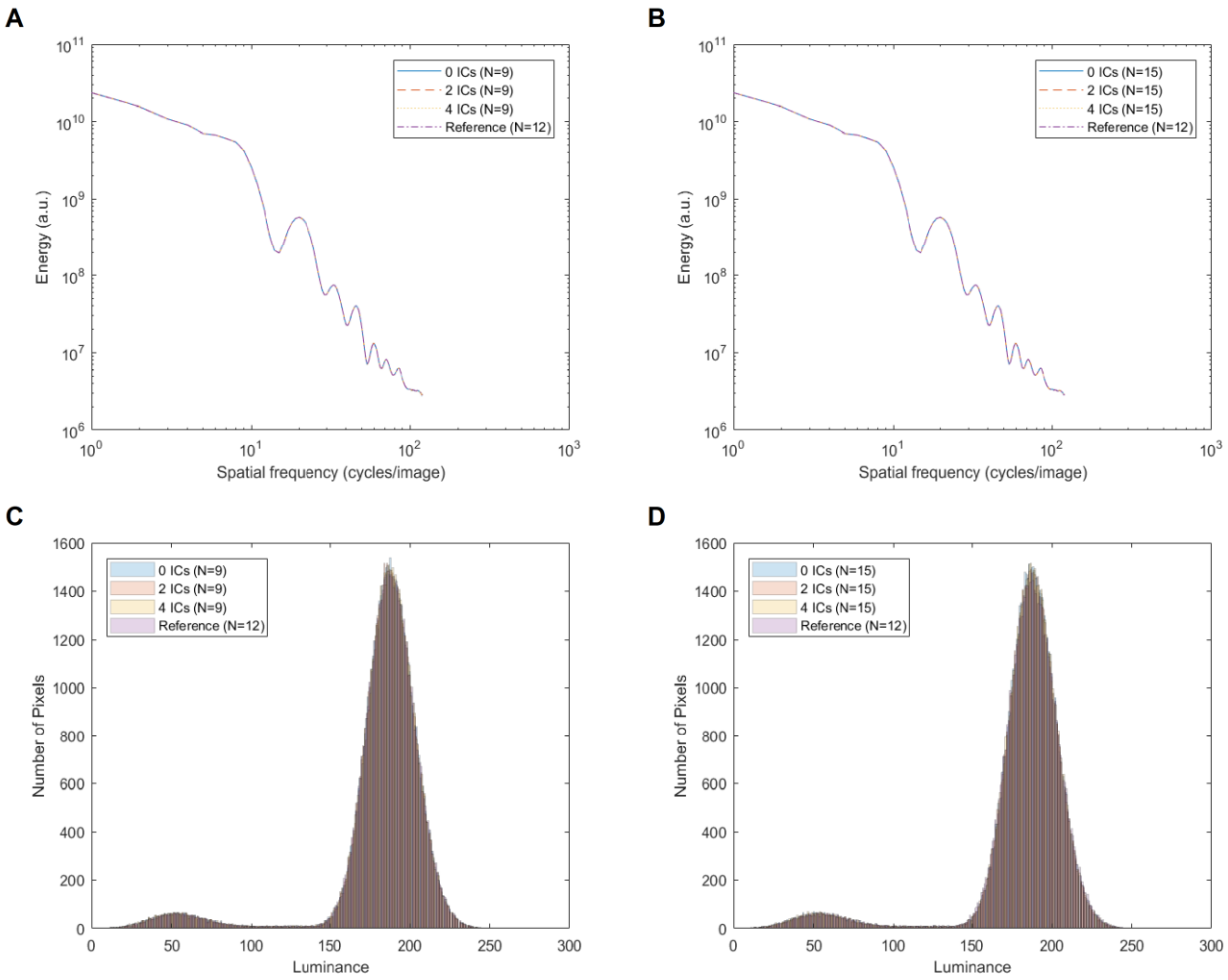


Figure 6.2: Rotational average of the Fourier energy spectrum (panel A and B) and luminance histogram profile (panel C and D) for two stimuli (test stimulus 9 and test stimulus 15) presented in the Experiment 1. Panel A and C show the low-level features statistics for the test stimuli "9" and 0, 2 or 4 ICs compared to the Reference (12 items) whereas panel B and D show the low-level features statistics for the test stimuli with 15 items and 0, 2 or 4 ICs compared to the Reference (12 items). Note that in all figures the curve profiles (0,2, 4 ICs and Reference) almost fully overlap, thus indicating an extremely high equalization of the low-level statistical properties of the stimuli.

6.3.2. Procedure

The experiment was conducted in a quiet and dimly illuminated room. Subjects were comfortably seated at a viewing distance of 80 cm from the screen. The general procedure was explained by the experimenter to each participant before starting the experiment and detailed instructions were also provided on the display. No information about the illusions or the connectedness was given to the subject. The participants performed a two-alternative forced choice task (2AFC) in which they were asked to choose the set containing more objects between two rapidly presented visual patterns by pressing the corresponding keys on the keyboard. The experimental phase was preceded by a brief training composed of 24 trials to allow the subject to familiarize with the task. In the training phase, we presented only 12 objects reference stimulus vs. 9 objects test stimulus (8 trials for each test set with 0, 2 or 4 ICs). Each experimental trial started with a black background (RGB = -1, -1, -1) lasting 1000 ms, followed by a fixation cross (Font: Times; Size: 16 pixels; RGB = 0, 0, 0) projected for 1000 ms alone, and then, two panels appeared at the left or right of the fixation cross (72 pixels between the nearest edge of each square panel to the fixation) for additional 800 ms (Figure 6.3). The side of the reference and test patterns was counterbalanced and randomized across trials. After the stimuli offset, an empty screen (RGB = -1, -1, -1) was presented until the participant's answer. The subjects could select the stimulus by pressing the appropriate key with their left or right index finger ("F" key for the left stimulus and "J" key for the right stimulus).

Response time was not restricted, but we emphasized in the instructions to answer as fast as possible. After the practice session, two counterbalanced blocks of 336 randomly ordered trials were presented, for a total of 672 experimental trials (32 trials for each of the 21 conditions), separated by a forced pause (3 min) at the half of the whole session and by 2 short forced pauses (1 min) at the half of each block. The whole experiment lasted around 40-45 min.

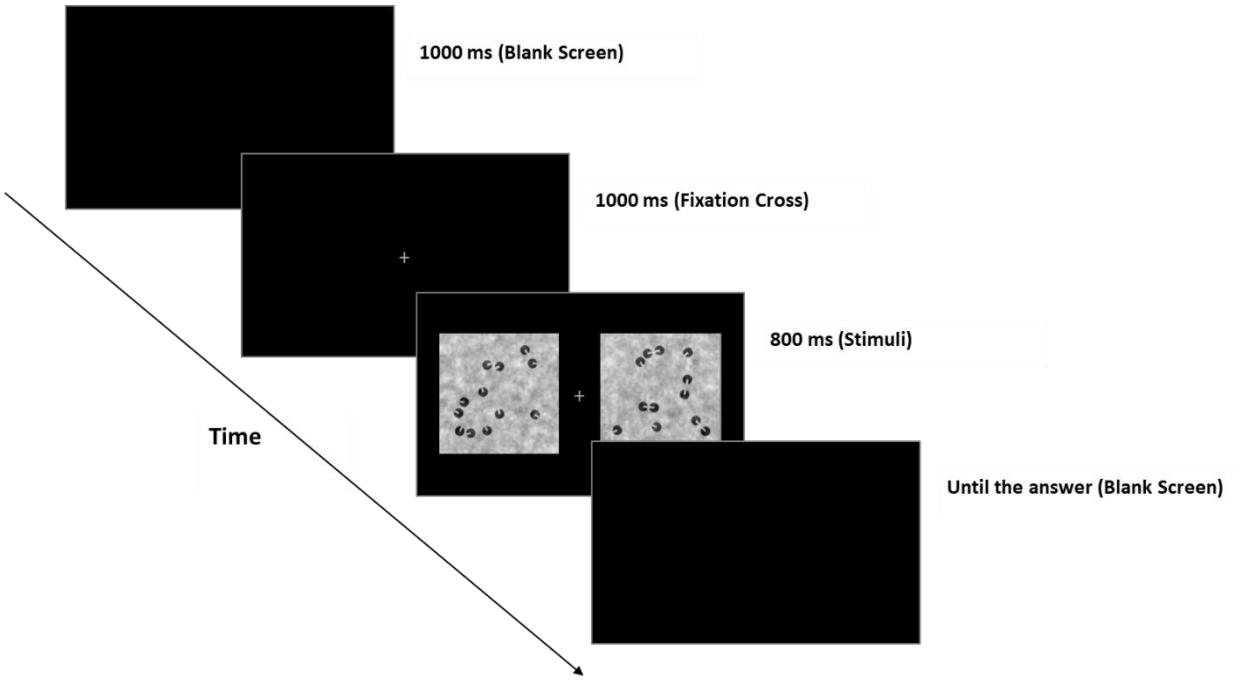


Figure 6.3: The numerical discrimination task. Subjects had to indicate the numerically larger between two stimuli by pressing the corresponding left or right key (F or J key). The side of reference and test pattern was balanced and randomized.

6.3.3. Statistical methods

The data were analyzed with *R-Studio* (2018, v. 3.6.2; http://www.rstudio.com/) and *Jamovi* (2019, v. 1.1.5; https://www.jamovi.org) softwares as in the Experiment 1, *Chapter 3.*

6.3.2. Results and Discussion of Experiment 1

Subjects performed a non-symbolic comparison task between 2 stimuli (e.g., which stimulus is numerically larger?). We manipulated the numerical distance between the reference (always 12 items) and the test (9 to 15 items) stimulus and the number of illusory connections (0, 2 or 4 ICs) in the image. All the stimuli were matched for SF content. Choice proportion and reaction times (RTs) were recorded.

Two separated one-way repeated measures analyses of variance (ANOVAs) with the number of ICs (0, 2, or 4) as within-subject factor were performed over both the individual PSEs and the CoV.

As can be visually observed in the figure plotting the psychometric functions obtained pooling over the aggregate data of the subjects (Figure 6.4A), the curved functions suggest that the overall performance was modulated by the numerical distance between test and reference. Furthermore, the rightward shift of the psychometric curves for 2 and 4 ICs conditions compared to the 0 ICs suggests an underestimation of the perceived numerosity as the number of connections in test stimuli increased.

In particular, we found a significant effect of the number of ICs over the PSEs, $F(2, 34) = 10.0$, $p < .001$, $\eta^2_p = .371$. That is, the PSEs increased with the number of ICs (Figure 6.4B). Such a pattern, was confirmed statistically by means of post-hoc comparisons (Bonferroni correction), revealing a significant difference between 0 and 2 ICs, $t(34) = -3.25$, $p < .01$, and between 0 and 4 ICs, $t(34) = -4.29$, $p < .001$. No difference was found between 2 and 4 ICs, $t(34) = -1.04$, $p = .92$. A polynomial trend analysis shows a significant positive linear trend only, $t = 4.29$, $p < .001$. In other words, more items in the test stimulus were required to compensate for the reduction in perceived numerosity due to the connected pairs (e.g., numerosity underestimation). This suggests that numerosity might be based on discrete object processing as showed in previous studies (e.g., Adriano, Rinaldi, & Girelli, under review; Kirjakovski & Matsumoto, 2016) since continuous features were equals across ICs levels. In addition, the analysis of CoV showed no significant effect of the number of ICs, $F(2, 34) = 0.53$, $p = .59$, $\eta^2_p = .030$, (Figure 6.4C), suggesting an equal numerical estimation precision across ICs conditions as predicted by the Weber's law.

We also ran a separate analysis of reaction times in the forced choice task. Data were logarithmically transformed and responses whose latencies fell outside of 1.5 times the interquartile range of the distribution were discarded (6.48% of the data).

A 3 x 4 repeated measures ANOVA with the number of ICs (0, 2, 4) and the absolute numerical distance between reference and test stimuli (0, 1, 2, 3) as within-subject factors was performed on the mean RTs data. A significant effect of the absolute numerical distance was found, $F(3, 51) = 24.326$, $p < .001$, $\eta^2_p = .59$. Post-hoc comparisons (Bonferroni correction) showed that except for the distance 0 from 1, $t(51) = -0.929$, $p = .99$, and for distance 2 from 3, $t(51) = 2.439$, $p = .110$, a significant difference between was found among all the other numerical distances ($p < .001$). A polynomial trend analysis shows a strong decreasing linear trend, $t = -7.75$, $p < .001$, but also a significant quadratic, $t = -2.38$, $p = .021$, and cubic trend, $t = 2.683$, $p = .010$, (Figure 6.4D).

Neither the main effect of the number of ICs, $F(2, 34) = 0.015$, $\varepsilon = .73$, $p = .96$, $\eta^2_p = .001$, nor the interaction between the number of ICs and the absolute numerical distance, $F(6, 102) = 1.00$, $\varepsilon = .57$, $p = .40$, $\eta^2_p = .056$, were significant.

Thus, besides the choice proportion we found that also the speed in discriminations was modulated, at least, by the numerical distance between stimuli (e.g., Buckley & Gillman, 1974), although power spectrum and spatial frequency were uncorrelated with the numerosity itself (Dakin et al., 2011). Indeed, RTs decreased as the numerical distance between reference and test increased, as suggested by the Weber's law. We did not find any effect of ICs over RTs, suggesting that independently of the number of connections, stimuli were processed at the same speed.

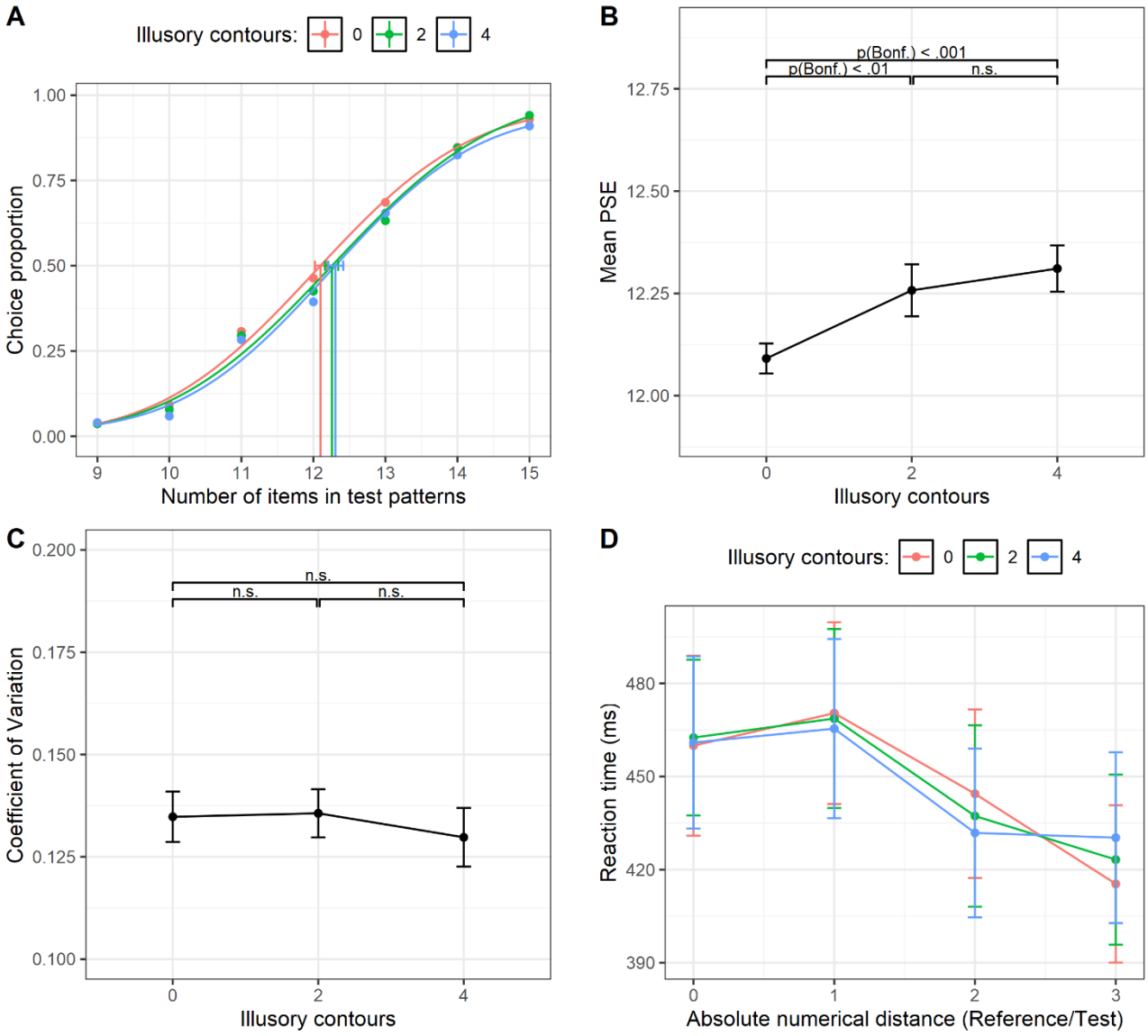


Figure 6.4: (A) Psychometric functions obtained fitting Gaussian cumulative distribution function (for 0-2-4 ICs) pooling over the aggregate data of all the subjects. Please note that this graph is reported to illustrate the technic, but all subsequent analysis was done with similar functions over individual subjects. The x-axis represents the actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were judged as more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each condition. The error bars represent the bootstrap 95% confidence intervals. (B) mean PSE in function of the number of illusory contours. (C) mean CoV in function of the number of illusory contours. D) mean RTs (inversely transformed) in function of the absolute numerical difference between refence and test stimuli and the number of illusory contours. The error bars represent ± 1 standard error of the mean (SEM).

6.4. Experiment 2: Estimation Task

To further extend the results of the first experiment, in the Experiment 2 we tested whether *scalar variability,* the fact that standard deviation of estimates increases proportionally to the numerosity resulting in a constant CoV (e.g., Whalen et al., 1999), is also preserved in the estimation task, when spatial frequency content of numerical stimuli is controlled. As in the first experiment, we also tested whether the perceived numerosity is affected by the degree of object segmentation (e.g., Adriano, Rinaldi, & Girelli, under review; Kirjakovski & Matsumoto, 2016). Subject were asked to report the number of items in the stimulus while we manipulated the numerosity (9 to 15 items) and the number of ICs (0, 2 or 4) in the sets. Importantly, all the stimuli were matched for SF content. According to Dakin's model, we expect that neither the presence of the scalar variability nor an effect of the ICs should be found in the estimation task.

6.4.1. Materials and methods

6.4.1.1. Participants

A new sample of 20[5] undergraduate and postgraduate students (*M* age = 24.8, *SD* = 9.68 years; 14 females, 19 right-handed), with normal or correct-to-normal vision, was recruited for the second study. All the subjects were naïve regarding the purpose of the experiment.

6.4.1.2. Stimuli and procedure

The stimuli used were the test patterns from Experiment 1. Participants were required to explicitly estimate the number of objects in the stimuli. Task instructions were provided orally by the experimenter and were also presented in the written format on the display. No information was given regarding the range of numerosity or the presence of illusions. Each trial began with a black background (RGB = -1, -1, -1) lasting 1000 ms, followed by a central fixation cross (Font: Times; Size: 16 pixels; RGB = 0, 0, 0) which was kept on the screen for 1000 ms, and then a stimulus appeared centrally for 800 ms. After the stimulus offset, a message (*"Estimation:"*; Font: Arial; Size: 30 pixels; RGB = 1, 1, 1) appeared centrally on the screen signaling that estimation was allowed. Then, the subject had to type the answer over a numerical keypad of a standard keyboard and had to press

[5] Another participant was tested but discarded from the analysis for unreliable performance (e.g., i.e., about 50% of trials answered with "0").

the space bar to confirm the response (Figure 6.5). The typed answer appeared on-line on the screen and subjects were allowed to modify it before confirming the input. After the space bar was pressed another trial began. The experimental phase was preceded by a brief training composed of 12 trials (without feedback) in which only the stimuli with 9, 11, 13, 15 items and 0, 2 or 4 ICs were projected. After the training phase, two counterbalanced blocks of 168 trials (336 total trials) presented in random order were projected. The blocks were separated by a forced pause of 3 minutes and the whole experiment took approximately 45 minutes.

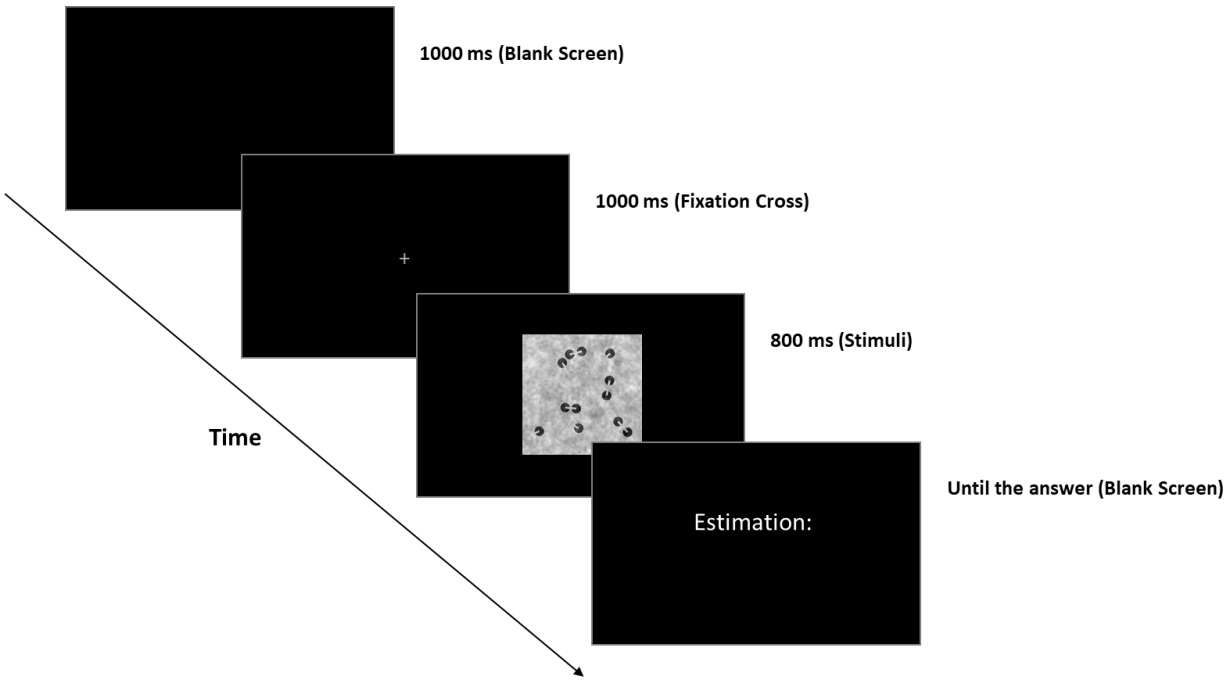


Figure 6.5: The numerical estimation task. Subjects had to estimate the number of objects contained in the set, by typing the answer on the numerical keypad of a normal pc keyboard.

6.4.2. Results and Discussion of Experiment 2

Subject had to report the number of items in the stimulus while we manipulated the numerosity (9 to 15 items) and the number of ICs (0, 2 or 4) in the sets. Stimuli were fully equalized for SF content. Subjective mean estimations were recorded. Data were inspected for the presence of possible outliers. We discarded trials in which no answer was given (N = 7) and trials that were below or above 2 SD from the grand mean (N = 35). Less than 1% of the data were discarded.

Data were thus analyzed with a 3 x 7 repeated measures ANOVA, with the number of ICs (0, 2, or 4) and the target numerosity (9, 10, 11, 12, 13, 14, 15) as within-subject factors and the mean subjective estimations as dependent variable.

We found a significant main effect of the target numerosity, $F(6, 114) = 215.61$, $\varepsilon = .24$, $p < .001$, $\eta^2_p = .91$, (Figure 6.6A). All post hoc comparisons (Bonferroni correction) were significant, indicating that all mean estimations for each target numerosity were significantly different from each other (all $p < .05$). A polynomial trend analysis shows a significant linear trend, $t = 35.82$, $p < .001$, and a significant quadratic trend, $t = -2.74$, $p < .01$.

A significant effect of the number of ICs was also found, $F(2, 38) = 4.88$, $\varepsilon = .66$, $p = .027$, $\eta^2_p = .205$, (Figure 6.6B). Post hoc comparisons (Bonferroni correction) showed a significant difference between 0 ICs and 4 ICs, $t(38) = 2.86$, $p = .02$. The difference between 0 ICs and 2 ICs was marginally significant, $t(38) = 2.50$, $p = .05$, whereas it was not between 2 ICs and 4 ICs, $t(38) = .361$, $p = 1$. A polynomial trend analysis shows a significant linear trend only, $t = -2.86$, $p < .01$.

Finally, we found a significant interaction between the target numerosity and the number of ICs, $F(12, 228) = 2.96$, $\varepsilon = .49$, $p = .01$, $\eta^2_p = .135$, (Figure 6.6C). Post hoc comparisons (Bonferroni correction) over each numerosity showed that only for the target numerosity “11” and “12” mean estimations decreased when ICs were manipulated. For the target numerosity “11” we found a significant difference only between 0 ICs and 4 ICs, $t(38) = 2.63$, $p = .03$, whereas for the target numerosity “12” we found a significant difference between 0 ICs and 4 ICs, $t(38) = 4.63$, $p < .001$, and between 2 “ICs” and 4 “ICs”, $t(38) = 3.74$, $p = .002$. All the other comparisons were not statistically significant (all $p > .05$).

This pattern of interaction suggests that the underestimation effect driven by the ICs was greater for the central values of the target numerosity range. One possible explanation could be that a degraded digital visual quality of the stimuli was present toward the extremities of the target numerosities as a result of the spatial frequency equalization (Willenbockel et al., 2010), which could

have weakened the perception of ICs over these numerosities. Central values may have a better digital quality because the average power spectrum (applied to each stimulus) was more similar to the original spectrum of these stimuli (combined with the original phase). In any case, the power spectrum was matched across ICs conditions over the whole numerical range. Since previous studies employing the same task and the same numerical range found a clear effect of ICs (but no interaction with numerosity) when the estimation was performed over the original stimuli, which did not have any image quality degradation toward the extremes (Adriano et al., under review; Kirjakovski & Matsumoto, 2016), we are confident that the ICs-underestimation effect is reliable.

We also ran a 3 x 7 ANOVA with target numerosity and number of ICs as within-subject factors and the CoV as dependent variable. Results showed that both the main effect of the number of ICs, $F(2, 38) = .05$, $p = .92$, $\eta^2_p = .003$, and the interaction between target numerosity and number of ICs, $F(12, 228) = .23$, $\varepsilon = .49$, $p = .96$, $\eta^2_p = .012$, were not statistically significant. However, the main effect of target numerosity was significant, $F(6, 114) = 4.64$, $\varepsilon = .51$, $p = .005$, $\eta^2_p = .197$. Post hoc comparisons (Bonferroni correction) showed a significant difference for the CoV of the target numerosity “9” compared to the numerosity “11”, $t(114) = -3.12$, $p = .047$, the numerosity “12”, $t(114) = -4.01$, $p = .002$, the numerosity “13”, $t(114) = -4.22$, $p = .001$, the numerosity “14”, $t(114) = -4.11$, $p = .002$, and the numerosity “15”, $t(114) = -3.86$, $p = .004$. Except for the numerosity “9”, all the other comparisons among numerosities from 10 to 15 were not statistically significant (all $p > .05$) and hence the CoV was constant across numerosities and ICs conditions (Figure 6.6D). Thus, numerosity estimations showed a scalar variability as suggested by psychophysical models of numerosity representation (e.g., Whalen et al., 1999), although power spectrum content was not predictive of numerosity. Concerning the smaller CoV for the target numerosity “9”, we note that a similar pattern has been recently reported in the literature for a similar numerosity (DeWind, Bonner, & Brannon, 2020). A possible explanation may be that a rapid counting combined with true approximation would be used for relatively small numerosities (Mandler & Shebo, 1982).

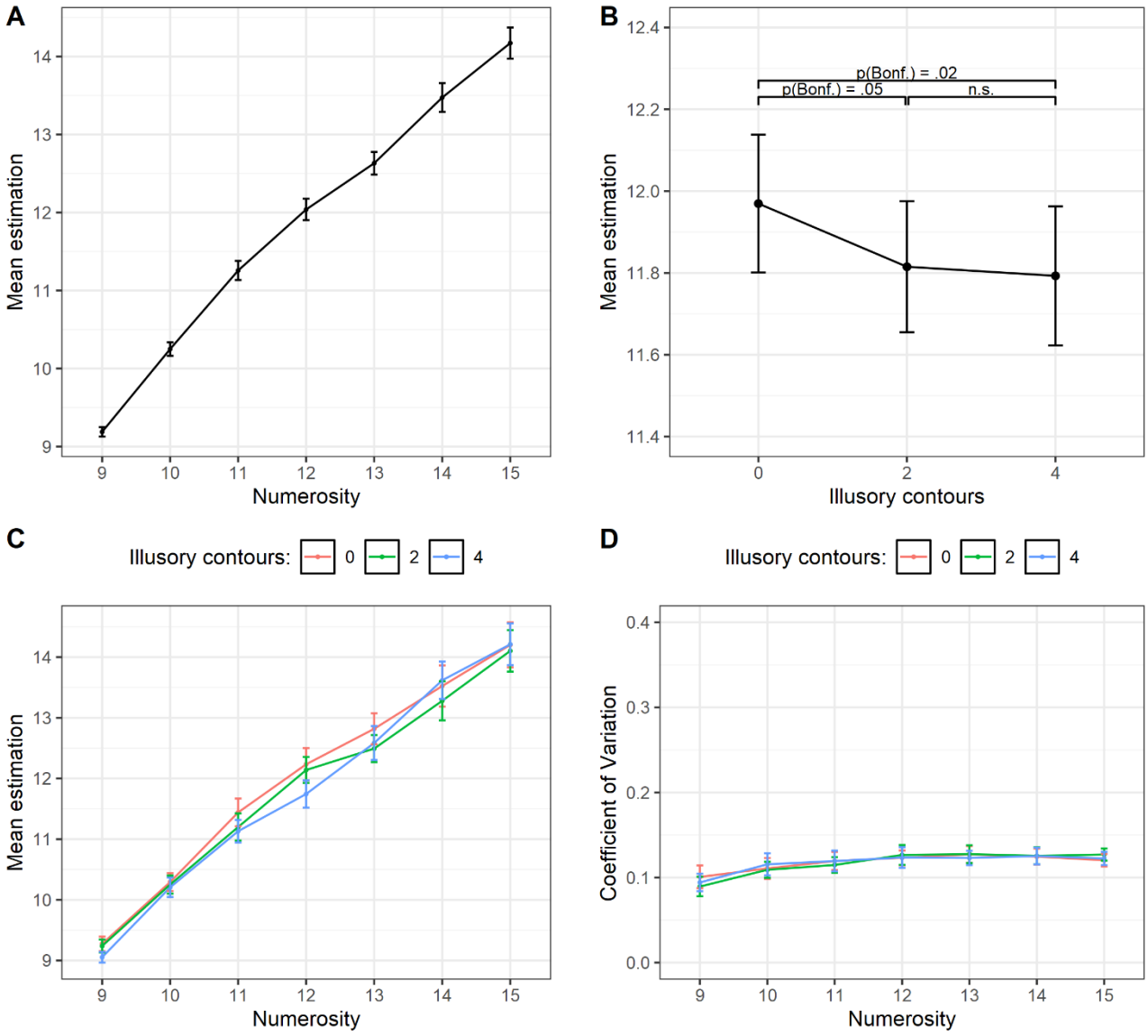


Figure 6.6: (A) Mean estimations as a function of the target Numerosity. (B) Mean estimations as a function of the number of ICs. (C) Mean estimations as a function of the target Numerosity and the number of ICs. (D) CoV as a function of the target Numerosity and of the number of ICs. The error bars represent ± 1 SEM.

6.5. General Discussion

In this study we investigated whether behavioral performance in numerical tasks follows the Weber's law even though power spectrum and luminance of the stimuli were kept uninformative about the numerosity in the stimuli and, thus, whether the ANS may exploit the number of segmented objects in the visual input (e.g., Dehaene & Changeux, 1993; Burr & Ross, 2008a).

The results of two behavioral experiments, using numerical comparison (Experiment 1) and estimation tasks (Experiment 2), strongly suggest that visual numerosity processing does not depend on the physical properties of the raw visual input such as spatial frequencies amplitude, texture luminance statistics or on other low-level visual cues (e.g., density and convex hull).

Indeed, we found that the participants' performance followed the Weber's law in both experiments, although we neutralized spatial frequencies amplitude and luminance histograms across the numerosities depicted in the stimuli. In particular, our results showed that choice proportion and reaction times in Experiment 1 were modulated by the numerical distance between test and reference sets. This is further confirmed by results from the Experiment 2, where we found that mean numerosity estimation (and SD) increased with the target numerosity resulting in a stable CoV across the range, despite luminance and power spectrum were constant in the whole set of stimuli. This pattern of results is in accordance with psychophysical models postulating that the encoding of non-symbolic numerosities follows the Weber's law (e.g., Whalen et al., 1999).

Crucially, when objects were connected by means of IC lines, the numerosity was underestimated in both experiments (although in Experiment 2 this effect was found only for the central values of the numerical range), even though low-level confounds were controlled across connectedness levels.

Together, these results indicate that visual segmentation of the single (perceptual) objects of an organized scene may have a key role for visual numerosity computation. The present findings are thus in line with recent studies suggesting that, at least for a moderate numerosity array, visual numerosity is extracted independently from low-level features, such as the raw spatial frequency of the image, using a segmented visual input (Adriano et al., under review; Anobile et al., 2017; Burr & Ross, 2008a; Kirjakovski & Matsumoto, 2016).

Interestingly, this is not the first time that texture models fail to predict the performance in numerical tasks. In a recent work, Anobile and colleagues (Anobile at al., 2017), in line with texture-density models, predicted that adding physical lines in the stimulus, which increases the energy in

the high spatial frequencies (and the physical density) of the array, should make the array appearing numerically larger than an equal array of isolated dots. However, as they showed, adding irrelevant lines that connected few dots actually increased the energy in the high spatial frequencies (and the physical density) of the stimuli, but the manipulation led to a decrease rather than to an increase of both the perceived numerosity and the perceived density. These results clearly contradicted the model of Dakin et al. (2011), but only over a moderate numerosity range (e.g., 15 items). Indeed, when the same connectedness manipulation was applied to a collection of 100 dots, the numerosity/density was overestimated following the predictions of the Dakin's energy model, thus suggesting that texture-density mechanisms may be at play for very high cluttered arrays.

In a further work, Cicchini and co-workers (Cicchini, Anobile, & Burr, 2016) tested the subjects in an "odd-one-out" task in which two standards numerical stimuli (with equal density and area) were presented with a deviant stimulus (the odd-one) varying in either area or density. The task was to select the stimulus differing from the other two, without knowing which dimension was changing. They showed that at low item-density, subjects were more sensitive to numerosity variation than either area or density but for very high numerosity/density, area and density were sensed directly and numerosity was calculated as a combination of these continuous features.

Furthermore, these studies reported that the participants' Weber fraction was constant over a low numerosity range, but, after a critical numerosity, it decreased proportionally to the square-root of numbers (see also Anobile, Cicchini, & Burr, 2014, 2016), suggesting the presence of a separate perceptual mechanisms for very high numerosities, perhaps based on texture-density.

These results are thus in line with recent proposals suggesting that collections of objects may be processed by means of 3 different visual mechanisms: one for very small numerosities (e.g., 1-4 items, or subitizing), one for moderate arrays (e.g., > 4 items, ANS) and one for higher arrays (e.g., ≈ 100 items, texture-density), where items are too cluttered or crowded and object segmentation becomes harder (Anobile et al., 2014, 2016; Kirjakovski & Matsumoto, 2016; Pomè, Anobile, Cicchini, Scabia, & Burr, 2019). Thus, three different systems may be recruited for numerosity processing, but perhaps contextual factors (e.g., the number of items in the scene) or task constraints (e.g., stimulus exposure time) could activate one of the three systems or promote different strategies based on the discrete numerosity or rather on the low-level visual cues (e.g., Dietrich, Nuerk, Klein, Moeller, & Huber, 2019).

Yet, one could argue that since phase information from the original stimuli was preserved, and thus the overall structure of the individual objects was not critically deteriorated (e.g., Joubert,

Rousselet, Fabre-Thorpe, & Fize, 2009), numerosity could be indirectly computed, for example, exploiting the density or the convex hull of the collections. Thus, these low-level visual features could be still used as proxy for the numerosity. Here, to discard further explanations based on density or convex-hull processing, we used the connectedness effect triggered by ICs lines to directly manipulate the level of object segmentation (Adriano et al., under review; Kirjakovski & Matsumoto, 2016). By means of this manipulation, we show that alternative explanations based on low-level cues (e.g., density, convex hull) should be rejected because our stimuli had the same convex hull (e.g., extent and shape), density and spatial frequency across connectedness levels. The connectedness effect represents thus a further striking evidence that numerosity processing exploits visual segmentation rather than simply global texture statistics of an unsegmented raw image. That is, visual processing of moderate numerosity may exploit a perceptually organized visual input, and the pairs of connected dots may be processed as perceptual input-units for numerosity processing (Adriano et al., under review; Franconeri et al., 2009; He et al., 2009; He et al., 2015; Kirjakovski & Matsumoto, 2016).

Recent computational models suggest that topological invariance of numerosity, as suggested also by the fact that numerosity perception is influenced by topological invariants such as the item connectedness, might be a key visual attribute distinguishing visual numerosity from continuous low-level features (He et al., 2015; Kluth & Zetzsche, 2016) and that the ANS may have naturally emerged from visual circuits originally dedicated to object perception, likely exploiting biologically plausible computational elements already existing in the visual cortex (DeWind, 2019; Kluth & Zetzsche, 2016; Nasr, Viswanathan, & Nieder, 2019). A final speculation could be that visual numerosity processing has redeployed basic neural mechanisms evolved to extract information for object perception independently of their shape or size, to summarize or track the total number of potential visual targets in the scene (Anderson, 2010).

In conclusion, our study suggests that power spectrum *alone* cannot explain the classic behavioral signatures of the ANS in non-symbolic numerical tasks (e.g., distance effect, scalar variability). Therefore, the Weber-like encoding of numerosity that we found in both experiments may reflect information that is not contained in the power spectrum, since this information was equalized across stimuli and hence kept uninformative about numerosity (Dakin et al., 2011; Durgin, 2008). Finally, the numerosity underestimation found when ICs were manipulated rather suggests that visual numerosity processing exploits a discrete, organized, visual input. Hence, we suggest that a discrete numerosity representation might arise following Gestalt principles of perceptual

organization that govern the figure/ground segmentation and perceptual unit formation, rather than from raw unorganized summary texture statistics.

Future studies should apply the spatial frequency equalization procedure used in this work, to test whether also numerosity-adaptation aftereffect (Burr & Ross, 2008a) is still present when subjects are adapted, and then tested, with numerical stimuli matched for spatial frequency content.

6.6. Conclusions

In sum, the present study demonstrates that numerosity processing is not merely based on low-level visual features such as the power spectrum of the image. Indeed, we found that a Weber-like behavior (e.g., numerical distance and scalar variability) was preserved even though power spectrum was uninformative about numerosity. Furthermore, numerosity was underestimated when objects were grouped by ICs lines. Taken together, the current work strengthens the notion that, at least for a moderate numerosity range, the ANS may exploit individual segmented objects of an organized visual scene, rather than mere summary texture statistics and global power spectrum *in its own.*

Chapter 7

7.1. The ratio effect in visual numerosity comparisons is preserved despite spatial frequency equalisation

Chapter adapted from: Adriano, A., Girelli, L., & Rinaldi, L. (2021). *Vision Research*, *183*, 41–52. https://doi.org/10.1016/j.visres.2021.01.011

7.2. Introduction

It has been proposed that numerosity processing would exploit general visual texture-density perception mechanisms, operating over an unsegmented raw visual input (Dakin et al., 2011; Durgin, 2008; Morgan et al., 2014; Tibber, Greenwood, & Dakin, 2012). For example, Durgin (2008) suggested that numerosity adaptation can be alternatively explained by an adaptation to low-level features such as texture-density (but see Burr & Ross, 2008b), or a correlated feature of density such as the statistical kurtosis of the visual image (e.g., how uniform is the grey level distribution in the image). Similarly, Dakin and colleagues (2011) proposed an influential model in which numerosity would be computed by "multiplying" the physical density by the field area occupied by the set. According to this model, numerosity (and density) would be indirectly derived from surrogate low-level information like the power spectrum of the stimulus; in other words, from the output of high SF and low SF tuned filters, operating over a raw visual image input. In such cases, high SF energy should be proportional to the number of objects (i.e., luminance edges and fine details), whereas low SFs should be related to the stimulus extent or field area (i.e., coarse details of the image). Similarly, Morgan and colleagues (2014) suggested that numerosity would be encoded by the amount of 'details' (e.g., edges) in the image, which would in turn be accurately captured by the high-spatial-frequency content. In sum, despite numerosity not being related to SF on a theoretical level (Burr & Ross, 2008a; Dehaene & Changeux, 1993), when dots are presented in a fixed region, their numerical value (and their perceived numerical distance) would be proportional to the number of light/dark cycles per image (e.g., SF) and to their amplitude (e.g., energy). For

instance, when two stimuli with 38 and 35 dots are presented in a fixed region, Fourier spectral decomposition of the stimulus with 38 dots will have higher power in the high SF range since it also contains more edges compared to the stimulus with 35 dots (for a similar graphical representation of the SF spectrum difference, see Anobile et al. 2017). This possibility has been supported by later studies indicating that observers often replace numerosity with SF (Anobile et al., 2017; Morgan et al., 2014). In essence, alternative proposals suggest that non-symbolic numerosity would be encoded by means of non-numerical mechanisms exploiting the integration or weighing of the raw low-level features confounded, or naturally correlated, with numerosity in the image (Gebuis et al., 2016; Leibovich et al., 2017).

Disentangling discrete numerosity information (e.g., the number of individual items) from continuous visual features confounded in the stimulus (e.g., SFs, luminance, density, etc) is thus a key empirical challenge in determining which visual mechanisms are at the core of the system deputed to numbers. Several methods have been proposed in the literature to control for low-level features correlated or confounded with numerosity (e.g., DeWind, Adams, Platt, & Brannon, 2015; Gebuis & Reynvoet, 2011; Piazza et al., 2004). For example, Gebuis and Reynvoet (2011) proposed a valuable method to generate numerical stimuli in which multiple visual features (e.g., convex hull, density, item size, cumulative surface and total perimeter) are not correlated with the numerosity across trials. However, none of these methods were able to wholly rule out the role played by low-level sensory information such as SFs. Indeed, the problem is not only that low-level cues could be correlated with numerosity across trials, but also that within each trial the available sensory cues could still be used as a basis for numerosity, especially in comparison tasks. This is an intrinsic problem concerning non-symbolic numerical cognition because it seems mathematically impossible to have full control over all the sensory cues in the stimuli (in the spatial domain). Indeed, two arrays of objects that differ for numerosity necessarily differ in at least one (or more) low-level features (Gebuis et al., 2016). Consequently, when we compare two numerical stimuli any difference in the available sensory cues within a given trial, which are inevitably also reflected in the raw power spectrum content of the image, may form the basis for numerosity judgments (Gebuis et al., 2016). In this sense, no matter the control applied to low-level visual features in the experiment, the generated stimuli will have an intrinsic SF content, and this information – as proposed by Dakin's model – would suffice to accomplish the discrimination task, even within each trial. Thus, numerosity judgments in the comparison task could theoretically be based merely on any sensory

cues (including SF information) differing between the stimuli, rather than on discrete numerosity information itself (Gebuis et al., 2016).

To date, to the best of our knowledge, only one study has carefully controlled for the possible effects of SF (and luminance) on non-symbolic numerical processing, making these features wholly uninformative of the target numerosity. In particular, Railo and colleagues (Railo, Karhu, Mast, Pesonen, & Koivisto, 2016; see also Adriano, Girelli, & Rinaldi, 2021) reported that the estimation time and error rate of natural human figures presented in a display increase with the target numerosity even when the SF content of the stimuli were fully equalised across all numerosities. Despite this, their study used a very limited numerical range (e.g., 1-6 items) in an explicit estimation task, and did not control for possible additional confounds such as low-level visual features correlated with numerosity across trials (e.g., density, convex hull, cumulative surface, etc.).

On these grounds, in the present study we systematically investigated whether non-symbolic number processing of a moderate range array (i.e., 8 to 18 items) operates over summary statistics of an unsegmented scene such as SF (Dakin et al., 2011; Durgin, 2008) or rather over the number of individual visual objects following Weber's law (Burr & Ross, 2008a; Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012). To this aim, we designed two experiments using the comparison task to test whether the behavioural signature of the ANS (ratio effect) is preserved despite the removal of SF as a cue for numerosity. In Experiment 1, we manipulated the ratio between the numerosity of the collections, generating experimental stimuli using Gebuis and Reynvoet's script (2011), which facilitates control of the main low-level features of the image (convex hull, density, item size, cumulative surface and total perimeter). The advantage of creating stimuli with this method is that no *single* visual cue is correlated with numerosity/ratios across trials. However, this does not preclude the possibility that numerosity can be simply extracted from the raw SF content of the image. Thus, all the stimuli used in the experiment were processed to equalise their Fourier amplitude spectra (e.g., SF, amplitude and orientations) and overall luminance statistics (e.g., mean, standard deviation, skewness, kurtosis), hence neutralising these low-level features cueing for numerosity (Railo et al., 2016). Using Fourier analysis, images were broken down into the sum of a set of sinusoidal gratings defined by four parameters: SF (e.g., dark/light cycles per image), amplitude (e.g., difference in luminance between the lightest and the darkest parts of bars), orientations (e.g., angle of the dark/light bars) and phase (e.g., position of the sinusoid relative to a reference point). Consequently, SF, amplitude and orientation parameters were equalised across

stimuli (keeping their original phase constant) so that numerosity or numerical distance between stimuli pairs cannot be inferred from these SF (or luminance) statistics (Willenbockel et al., 2010).

In Experiment 2, to further understand the effect of the power spectrum equalisation in numerical tasks, participants performed the same comparison task but were presented with both the original stimuli (i.e., not controlled for SF) generated using the method of Gebuis and Reynvoet (2011) and the SF equalised version of the same stimuli.

7.3. Experiment 1: Equalised Spatial Frequencies

In Experiment 1, we investigated whether the ratio effect is preserved over and above the SF content of the stimuli (e.g., Railo et al., 2016). Participants performed a comparison task in which they had to select the numerically larger of two sets of dots. We manipulated the numerical *ratio* (smaller/larger numerosity) between the stimuli across 3 symmetrical levels of increasing difficulties (ratios: 0.66, 0.75, 0.8). Crucially, all the generated stimuli were equalised for their SF content (e.g., SF, amplitude and orientations) and luminance across numerosities (and ratios). We predicted that if numerosity is merely based on the raw SF content of the image, since both stimuli have the same SF profile, no effect of the *numerical ratio* should be found (e.g., accuracy should be at chance level for each ratio). On the other hand, if numerosity encoding is not primarily based on simply the power spectrum, the discrimination should *depend* upon the *numerical ratio* between reference and test stimuli, as expected by a Weber-like encoding of numerosity (e.g., accuracy should decrease as the numerical ratio increases).

7.3.1. Materials and methods

7.3.1.1. Participants

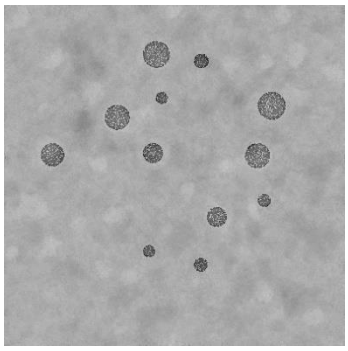
A sample of 28 participants (21 females, 26 right-handed) took part in the study. The mean age was 27.8 years (*SD* = 6.72). Due to Covid restrictions, all participants were recruited through a shared link from Pavlovia, a launch platform allowing researchers to conduct online PsychoPy experiments (www.pavlovia.org). All participants had normal or corrected-to-normal vision and were unaware of the purpose of the experiment. Each subject signed an online informed consent document before the experiment began and the study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Local Ethical Committee (protocol N° RM-2020-230).

7.3.1.2. Stimuli and Design

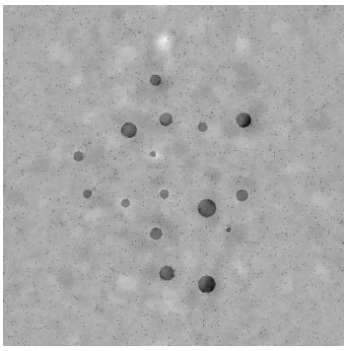
Original stimuli were generated off-line with Gebuis and Reynvoet's script (2011), which provides multiple controls over low-level visual cues (Figure 7.1). This script generates arrays of dots controlled for the 5 main continuous variables: area extended (or convex hull, the virtual elastic

enclosing the dots), total surface (the aggregate surface of all dots in one array), density (area extended/total surface), item size (average diameter of the dots presented in one array) and total circumference (circumference of all dots in one array, taken together). Post-hoc analyses ensured the absence of a relationship between numerical distance and the difference in visual properties (all R^2 values < .05; see also the Visual parameters analysis of the original stimuli section in the Supplementary Materials). Each stimulus was composed of black dots of a random size (RGB = 0, 0, 0) depicted on a middle grey background (RGB = 127, 127, 127) and scattered across a squared stimulus panel (395 x 395 px). A total of 192 stimuli with a different spatial pattern were generated (96 stimuli pairs). The numerosities depicted in the stimuli ranged between 8 and 18 dots. In each pair of stimuli, one set always contained 12 dots (reference), whereas the second numerosity (test) was smaller than 12 in half of the trials (8, 9 or 10 dots) and larger than 12 in the other half (14, 16 or 18 dots) resulting in three symmetrical ratios (smaller numerosity/larger numerosity: ratio 0.66, ratio 0.75 and ratio 0.8) around the reference numerosity (for a similar study with an analogous design and ratios, see Smets, Sasanguie, Szücs, & Reynvoet, 2015; Zhang, Wu, Wu, Mou, & Yue, 2020). Thus, 6 different comparisons between test and reference were generated in the stimuli: 8 vs 12 (ratio 0.66), 9 vs 12 (ratio 0.75), 10 vs 12 (ratio 0.8), 12 vs 14 (ratio 0.8), 12 vs 16 (ratio 0.75) and 12 vs 18 (ratio 0.66). For each of the 6 number comparisons, the script generated 16 pairs with different spatial patterns (half of the pairs with the relative numerically smaller set on the centre-left and the other half on the centre-right of the screen).

Following a similar methodology used in the *Chapter 6*, visual low-level statistical properties such as power spectrum and luminance histograms were fully equalised by processing all the experimental stimuli with the SHINE toolbox for MatLab (Willenbockel et al., 2010), As can be observed in Figures 7.2 and S1, indeed, stimuli have similar SFs amplitude spectrums and luminance profiles across numerosities and ratios.



Reference



Test Stimulus

Figure 7.1: The reference stimulus was always composed of 12 items. Test stimuli varied from 8 to 18 items across 3 numerical ratios. All the (original) stimuli were generated using the method of Gebuis and Reynvoet (2011). All the test and reference stimuli were equalised for SFs and luminance histogram profiles (See also Figures 2 and S1) using the SHINE toolbox (Willenbockel et al., 2010).

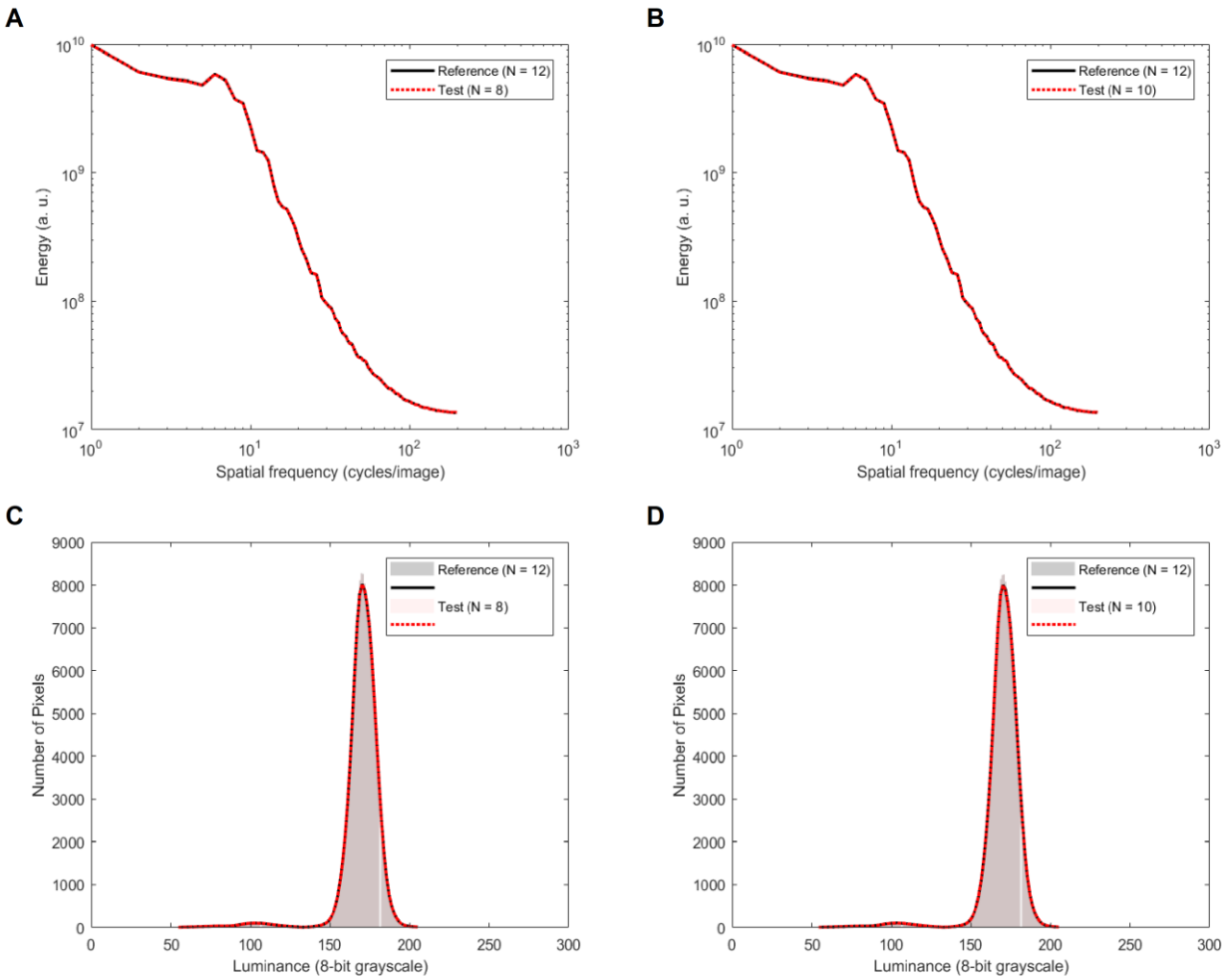


Figure 7.2: Rotational average of the Fourier energy spectrum (panels A and B) and luminance histogram profile (panels C and D) for two stimuli comparisons with different ratios (.66 and .8), as presented in Experiment 1. Panels A and C show the low-level feature statistics for the test stimuli with 8 items compared to the Reference (ratio 0.66), whereas panels B and D show the low-level feature statistics for the test stimuli with 10 items compared to the Reference (ratio 0.8). Note that in all figures the curve profiles almost fully overlap, thus indicating an extremely high equalisation of the low-level statistical properties of the stimuli. Stimuli images were coded in linear RGB 8-bit grayscale values.

7.3.1.3. Procedure

The stimuli were projected by means of an online PsychoPy routine (Peirce, 2007) and all the experimental materials (stimuli, etc.) were downloaded from Pavlovia repository to a temporary local folder stored on the computer of each participant prior to starting the experiment. The general procedure was explained to each participant by means of detailed instructions provided on the display before starting the experiment. The experimental task was a simple comparison between two rapidly presented arrays of dots (e.g., to determine which stimulus is numerically larger).

The experimental phase was preceded by a brief training period composed of 6 trials to allow the participant to familiarise themselves with the task. In the training phase, we presented only the condition with 12 items in reference vs. 18 items in test (ratio 0.66). Each trial started with a blank screen for 1000 ms (RGB = 0, 0, 0), before a grey fixation cross (Font: Times; Size: 16 pixels; RGB = 127, 127, 127) was presented for an additional 1000 ms. Next, two stimuli were simultaneously displayed within a black window (RGB = 0, 0, 0) to the left and right of the fixation cross (72 pixels between the nearest edge of each square panel to the fixation) for a maximum of 500 ms; afterwards, a blank screen (RGB = 0, 0, 0) was presented until the answer. Participants could respond either during the presentation of the stimuli (terminating the trial) or during the blank screen. They were instructed to press the left key ("F" key) with their left index finger if they judged the numerosity on the left side of the screen to be larger or to press the right key ("J" key) with their right index finger if they judged the numerosity on the right side to be larger (Figure 7.3). After the training phase, two equal counterbalanced blocks composed of 96 randomised trials were presented, for a total of 192 experimental trials. Within a block, each of the 6 comparison pairs (test vs reference) were repeated 16 times, resulting in 96 total trials per block (16 trials × 6 comparison pairs). The side of the reference and test patterns were counterbalanced and randomised across trials. The two blocks were separated by a self-paced pause and the experiment as a whole lasted approximately 11-13 minutes.

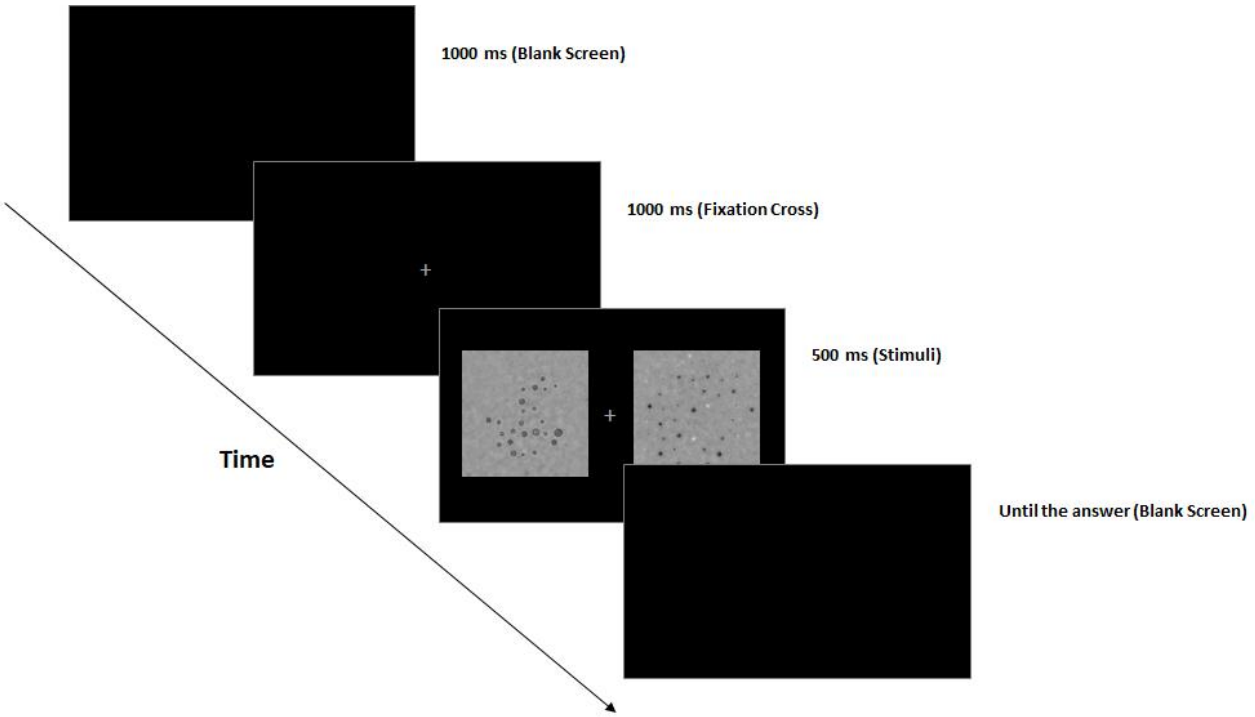


Figure 7.3: Stimuli were displayed for a maximum of 500 ms. The subject had to select the numerically larger stimulus. A total of 192 trials (96 trials × 2 blocks) were displayed.

7.3.2. Results

The data were analysed with *R-Studio* (2018, v. 3.6.2; http://www.rstudio.com/) and *Jamovi* (2019, v. 1.1.5; https://www.jamovi.org) softwares. Mean accuracy and reaction times (RTs) were calculated for each ratio and each participant (3 ratios: 0.66, 0.75 and 0.8). Accuracy and RTs of correct responses were analysed separately.

We ran a one way-repeated measures ANOVA with the numerical ratio (0.66, 0.75 and 0.8) as the within factor and the accuracy as the dependent variable. Results revealed a significant effect of the ratio, $F(2, 54) = 80.8$, $p < .001$, $\eta^2_p = .749$, suggesting that numerical discrimination becomes harder as the numerical ratio increases (Figure 7.4A). Post-hoc comparisons (Bonferroni correction) revealed a significant difference between ratios 0.66 and 0.75, $t(54) = 4.16$, $p < .001$, between ratios 0.66 and 0.8, $t(54) = 12.48$, $p < .001$ and between ratios 0.75 and 0.8, $t(54) = 8.32$, $p < .001$.

A similar analysis was conducted over RTs of correct responses only. Data were log-tranformed, and we discarded outliers, i.e., responses whose latencies fell outside 1.5 times the interquartile range of the data distribution. A total of 49 trials were discarded from the initial sample of 4223 observations (1.16% of the data). The results of the ANOVA showed a significant effect of

the ratio over RTs, $F(2, 54) = 6.47$, $p = .003$, $\eta^2_p = .19$. Indeed, discrimination becomes slower as the numerical ratio between numerosities increases (Figure 7.4B). Post-hoc comparisons (Bonferroni correction) determined the existence of a significant difference between ratios 0.66 and 0.8, $t(54) = -3.59$, $p = .002$, but not between ratios 0.66 and 0.75, $t(54) = -1.84$, $p = .21$, nor between ratios 0.75 and 0.8, $t(54) = -1.75$, $p = 0.25$.

As a further independent index of the numerical acuity, we also evaluated the Coefficient of Variation (CoV) of each subject, an index of the Weber fraction (e.g., Halberda & Odic, 2014). The psychometric function was generated by fitting the Gaussian cumulative distribution function to the data and parameters were estimated with a parametric approach based on the maximum likelihood method, using *Quickpsy* package for *R* (Linares & López-Moliner, 2016). The CoV (or Weber fraction) was computed as the ratio between the standard deviation (SD[6]) and the point of subjective equality (PSE) of the psychometric function (e.g., Helbig & Ernst, 2007). In order to minimise biases in estimating the parameters of the psychometric function, we fitted the psychometric curve taking into account the typical lapse in performance (e.g., missing a trial, finger-errors) by allowing the value of the guess rate (γ) and lapse rate (λ) parameters to vary in the default range of 0 – 0.05 (Wichmann & Hill, 2001). The 95% confidence interval of PSE was estimated by running 200 bootstrap data resamples (Figure 7.4C). Thus, for each subject we calculated the individual CoVs (e.g., SD/PSE) of the fitted psychometric functions (See Supplementary Materials for the individual psychometric functions and their respective goodness-of-fit; Figure S7.2 and Table S7.1). Hence, we tested the possible relationship between the individual precision (CoV) of the numerical representation and the individual overall mean accuracy score (e.g., the averaged accuracy score across ratios)[7] in the comparison task. We found a highly strong negative correlation between the two individual measures, $r = -.95$, $p < .001$. Indeed, as expected, subjects with better precision in the underlyng numerical representation (smaller CoV) achieved higher overall mean accuracy scores (Figure 7.4D).

[6] The JND (Just Noticeable Difference) is defined as the difference between the PSE and the 0.84 point on the psychometric function, i.e., the minimum increment in numerosity for which the test stimulus is judged to be more numerous 84% of the time). The 0.84 point is usually used to define the JND because then the JND corresponds to the SD of the cumulative Gaussian fitted to the data (e.g., Helbig & Ernst, 2007).

[7] This correlation analysis was performed over a sample of 27 subjects. One subject (n°15) was excluded after presenting an extremely large CoV (e.g., outlier). Note that all the previous results reported were also replicated; in doing so, they removed this participant from the analyses.

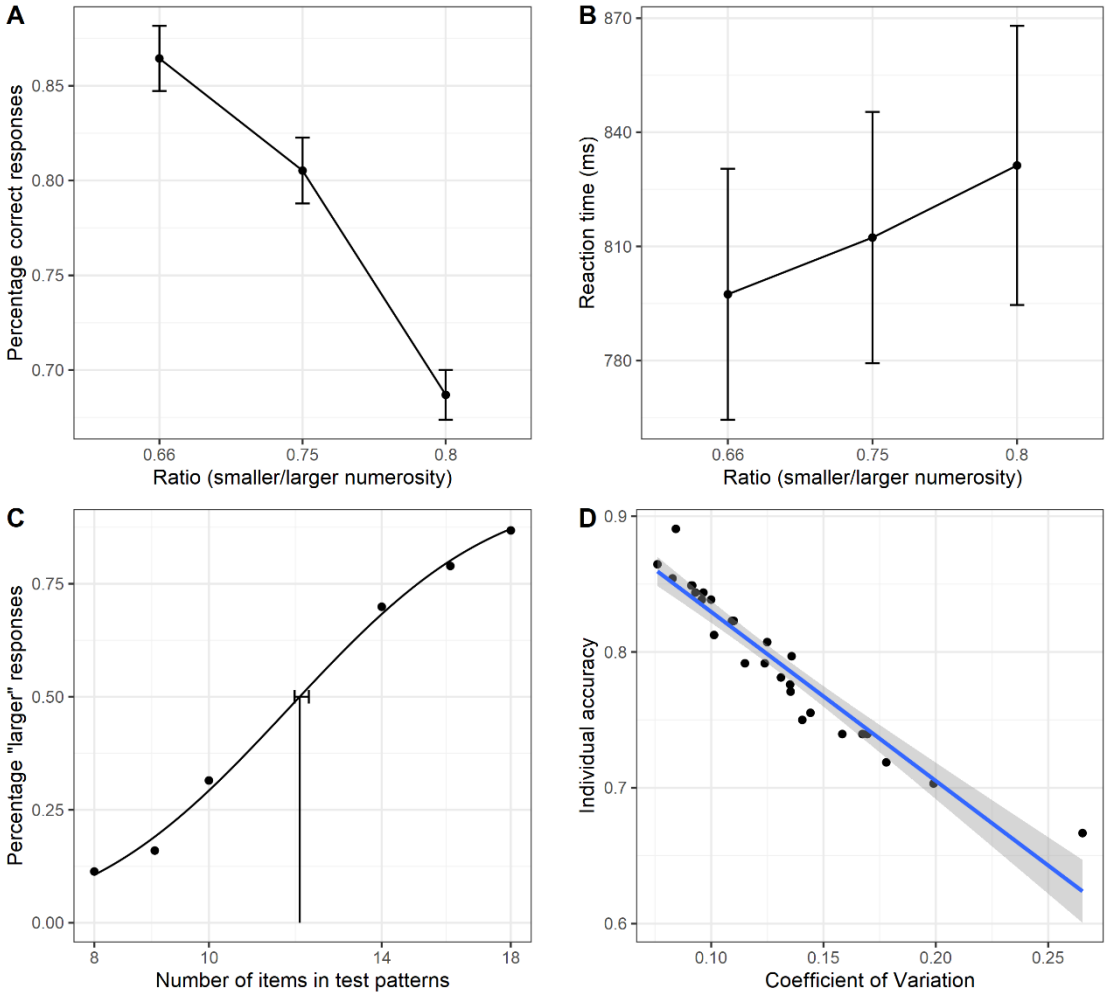


Figure 7.4: (A) Percentage of correct responses as a function of the numerical ratio. (B) Mean RTs of correct responses as a function of the numerical ratio. The error bars represent ± 1 standard error of the mean (SEM). (C) Psychometric function obtained fitting Gaussian cumulative distribution function pooling over the aggregate data of all the subjects. Please note that this graph is reported to illustrate the technic, but all subsequent analysis was done with similar functions over individual subjects. The x-axis represents the (log-transformed) actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were correctly judged as being more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each stimulus type. The error bars represent the bootstrap 95% confidence intervals (CI). (D) Scatterplot showing the correlation between individual CoV and overall mean accuracy scores. Shaded regions represent the 95% CI of the correlation line.

7.3.3. Discussion of Experiment 1

The results of Experiment 1 suggest that performance in the comparison task followed Weber's law. Indeed, contrary to the predictions of Dakin's model, we found that accuracy in discrimination decreased as the numerical ratio approached 1. Similarly, the RTs were slower when the numerical ratio increased. The fact that a clear ratio-based performance was identified despite stimuli SF content being entirely uninformative about numerosity suggests that the power spectrum may only have a partial role in numerosity processing.

In sum, we found that in comparison tasks the ratio effect was preserved, although the SF of the stimuli was equalised across numerosities/ratios (e.g., Railo et al., 2016). Thus, SF content alone cannot fully explain the emergence of the Weber-like performance (e.g., ratio effect) in non-symbolic numerosity tasks. In the next experiment, to better understand the role of SF and the effects of the removal of this information in numerical processing, we compared the performance in a comparison task with both original and SF equalised stimuli. For instance, it has been shown that in scene categorisation tasks the SF equalisation merely causes a drop in performance (e.g., slower RTs), but participants were still able to perform the task (Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Girard & Koenig-Robert, 2011).

7.4. Experiment 2: Original vs. Equalised Spatial Frequencies

To replicate and further extend the results of the previous experiment, in Experiment 2 we aimed to determine whether and to what extent the performance would be affected when the power spectrum is no longer informative about numerosity. We therefore compared the original set of stimuli with the set of equalised stimuli to establish a benchmark of the performance, adopting the same task used in Experiment 1. We hypothesise that the removal of the power spectrum information as a further sensory cue confounded with numerosity should merely cause a slight drop in the overall performance, as has been shown in the scene categorisation task (Joubert et al., 2009), but not a total disruption of the task. Hence, we predict that overall performance (e.g., accuracy and RTs) should be better with the original stimuli (e.g., in which the power spectrum is confounded/correlated with numerosity) compared to the equalised stimuli (e.g., in which the power spectrum is not confounded/correlated with numerosity). Therefore, if numerosity processing does not exploit the information contained in the power spectrum alone, the performance should decrease for higher ratios (ratio effect) but, crucially, this effect should be completely independent of the stimulus type used.

7.4.1. Materials and methods

7.4.1.1. Participants

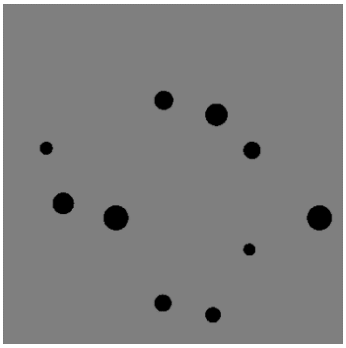
A new sample of 27[8] undergraduate and postgraduate students (*M* age = 28.44, *SD* = 4.97 years; 19 females, 25 right-handed), with normal or corrected-to-normal vision, was recruited for the second experiment. All the subjects were unaware of the purpose of the experiment.

7.4.1.2. Stimuli, Design and Procedure

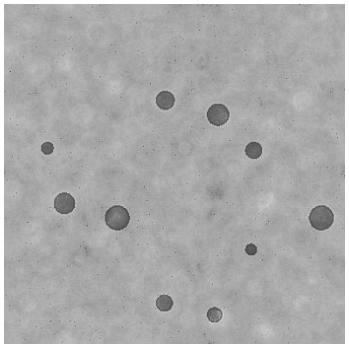
This experiment replicated the design and the procedure of Experiment 1. The only difference was that in each block, in half of the trials we presented the original stimuli (i.e., non-equalised) as extracted directly from the script of Gebuis and Reynvoet (2011), whereas in the other half we presented the same set of stimuli with the SFs equalised across numerosities, as in

[8] Another participant was tested but discarded from the analysis after providing an unreliable and extremely inaccurate performance (i.e., the subject did not understand the task or was not collaborative).

Experiment 1 (Figure 7.5). The experiment comprised two counterbalanced blocks of 96 random trials, separated by a self-paced pause.



Original Stimulus



SF equalized Stimulus

Figure 7.5: An example of the stimuli used in Experiment 2. Reference and test stimuli were presented in their original grey-level version, as generated by the script of Gebuis and Reynvoet (2011) and their respective version with the averaged SF spectrum processed with the SHINE toolbox (Willenbockel et al., 2010), as in Experiment 1.

7.4.2. Results

Similarly to Experiment 1, we calculated the mean accuracy and RTs for the two different types of stimuli for both each ratio and each participant. The accuracy and RTs of correct responses were separately analysed.

We first ran a repeated measures ANOVA (2 x 3) with the stimulus type (Equalised vs. Original) and the numerical ratio (0.66, 0.75, 0.8) as within-subject factors and the accuracy as a dependent measure (Figure 7.6A). We found a significant main effect of the numerical ratio on accuracy, $F(2, 52) = 86.7$, $p < .001$, $\eta^2_p = .76$. Post-hoc comparisons (Bonferroni correction) confirmed a significant difference between the ratios 0.66 and 0.75, $t(52) = 5.05$, $p < .001$, between the ratios 0.66 and 0.8, $t(52) = 13.06$, $p < .001$ and between the ratios 0.75 and 0.8, $t(52) = 8.01$, $p < .001$. Furthermore, we also found a significant main effect of the stimulus type, $F(1, 26) = 6.75$, $p = .015$, $\eta^2_p = .20$, suggesting there existed an overall higher accuracy for the original stimuli compared to the spatial frequency equalised stimuli. Crucially, no interaction between the two factors was found, $F(2, 52) = 0.17$, $p = .84$, $\eta^2_p = .007$.

We subsequently ran a similar analysis for the RTs of correct responses (Figure 7.6B). Before running the analysis, we log-transformed the data and discarded any responses whose latencies fell outside 1.5 times the interquartile range of the distribution. A total of 55 trials were discarded from the initial sample of 4049 observations (1.35% of the data). We found a significant main effect of the numerical ratio, $F(2, 52) = 7.15$, $p = .002$, $\eta^2_p = .21$. Post-hoc comparisons (Bonferroni correction) revealed a significant difference between the ratios 0.66 and 0.8, $t(52) = -3.65$, $p = .002$ and between the ratios 0.75 and 0.8, $t(52) = -2.67$, $p = .031$, but not between the ratios 0.66 and 0.75, $t(52) = -0.99$, $p = .98$. We also found a significant main effect of the stimulus type, $F(1, 26) = 11.32$, $p = .002$, $\eta^2_p = .30$, suggesting that RTs were slower with SF equalised stimuli as compared with the original stimuli. Finally, no significant interaction was found between ratio and stimulus type, $F(2, 52) = 0.29$, $p = .74$, $\eta^2_p = .011$.

As in the first experiment, we also evaluated the individual CoV, an index of the Weber fraction, for the two stimulus type conditions (e.g., Halberda & Odic, 2014). Following the procedure described in the first experiment, the CoV was computed as the ratio between the SD and the PSE of the psychometric functions for each stimulus type (Figure 7.6C). Thus, for each subject we calculated the CoVs (e.g., SD/PSE) of the fitted functions for the two stimulus types (See Supplementary Materials for the individual psychometric functions and their goodness-of-fit; Figure

S7.3 and Table S7.2, respectively). The mean CoVs for the two conditions were analysed with a paired samples t-test, revealing a significantly smaller CoV for the original stimuli compared to equalised stimuli, $t(26) = 2.61$, $p = .015$, $d = .50$, suggesting overall better acuity in numerical discrimination with the original stimuli (Figure 7.6D).

Finally, we also tested the possible relationship between individual CoVs in the two conditions (e.g., equalised vs. non-equalised stimuli). Results identified a significant correlation (Figure 7.6E) with a robust positive relationship between the two measures of acuity in each condition, $r = .68$, $p < .001$. In sum, subjects with better acuity in one condition also had better acuity in the other and vice versa, suggesting that a common sensory mechanism may drive the numerical discrimination performance over both types of stimuli.

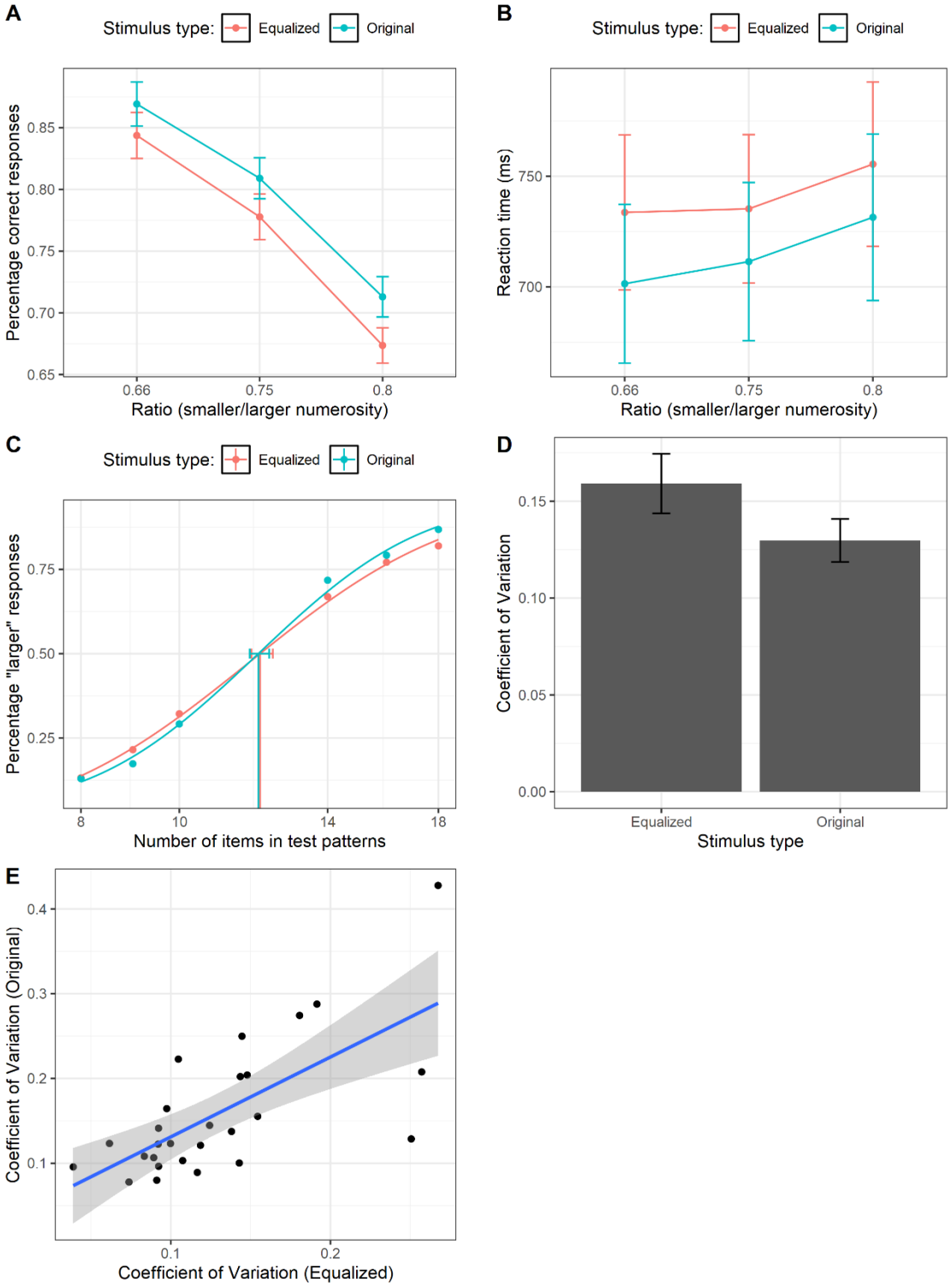


Figure 7.6: (A) Percentage of correct responses as a function of the numerical ratio and the stimulus type. (B) Mean RTs of correct responses as a function of the numerical ratio and the stimulus type. The error bars represent ± 1 SEM. (C) Psychometric functions obtained fitting Gaussian cumulative distribution function pooling over the aggregate data of all the subjects. Please note that this graph is reported to illustrate the technic, but all subsequent analysis was done with similar functions over individual subjects. The x-axis represents the (log-transformed) actual number of items in test patterns, whereas the y-axis shows the proportion of test patterns that were correctly judged as being more numerous than the reference. Vertical lines represent the PSE (0.5 threshold level) for each stimulus type. The error bars represent the bootstrap 95% CI. (D) CoV as a function of the stimulus type. (E) Scatterplot showing the correlation between conditions. Shaded regions represent the 95% CI of the correlation line.

7.4.3. Discussion of Experiment 2

As expected, in Experiment 2, we found that the overall performance was better with the original stimuli (i.e., higher accuracy and faster RTs) compared to their equalised counterparts (e.g., Joubert et al., 2009; Girard, & Koenig-Robert, 2011). In addition, we established that the CoV was also smaller (e.g., better precision) with the original stimuli. However, performance worsened when the numerical ratio increased, and crucially, we found that the ratio effect was independent of the type of stimulus, suggesting in turn that the ratio effect may emerge from information that is not primarily encoded in the power spectrum. This pattern of results suggests that the power spectrum may improve numerical precision but does not fully determine the Weber-like encoding of the numerosity itself.

Findings from Experiment 2 indicate that the power spectrum *per se* might not be the building block of numerosity and cannot explain the numerical ratio-based performance, despite numerical performance potentially being enhanced with the correct (i.e., original) amplitude spectrum (e.g., Joubert et al., 2009; Smets et al., 2015). It is not unlikely for the power spectrum to be considered and would simply act as many other sensory cues that are informative about numerosity. Indeed, prior research indicates that when several visual cues are correlated with numerosity, the performance in discrimination tasks is better (e.g., smaller CoV) compared to when fewer cues are correlated with numerosity (e.g., Gebuis & Reynvoet, 2012a; Smets et al., 2015). For example, Smets and colleagues (2015) showed that the performance is degraded (e.g., larger CoV) when multiple sensory cues are controlled for as compared to the condition with only a single sensory control or to the condition in which the visual cues were fully correlated with numerosity.

7.5. General Discussion

The present study investigated whether visual numerosity processing of a moderate array of items (e.g., 8-18 items) exploits summary texture statistics and/or the power spectrum confounded with numerosity (Dakin et al., 2011; Durgin, 2008; Gebuis et al., 2016) or rather a numerical representation of the individual objects in the set based on Weber's law (e.g., Dehaene & Changeux, 1993). To this aim, we tested whether behavioural performance in numerical comparison tasks follows Weber's law (ratio effect) even though stimuli power spectrum and luminance remained uninformative of numerosity (e.g., Railo et al., 2016).

The results of two experiments, adopting a numerical comparison task, strongly suggest that visual numerosity processing does not depend on the physical properties of the raw visual input, such as SFs amplitude or texture luminance statistics alone, at least in the tested range. Indeed, we discovered that participants' performance followed Weber's law in both experiments, although we neutralised SFs amplitude and luminance histograms across the numerosities depicted in the stimuli. In particular, our results showed that both accuracy and RTs were modulated by the numerical ratio between test and reference sets in both Experiments 1 and 2. Crucially, this was further corroborated by results from Experiment 2, where we found an analogous ratio effect in both original and equalised stimuli. Indeed, although the overall performance was slightly impaired with SF equalised stimuli (i.e., as compared to non-equalised stimuli), no interaction was found between the ratio and the type of stimuli, suggesting that the power spectrum may have a role in numerosity precision (e.g., CoV) but cannot explain the emergence of the ratio effect.

In sum, we found that the ratio effect in the comparison tasks was preserved although the power spectrum remained uninformative of numerosity. This pattern of results is in accordance with psychophysical models postulating that non-symbolic numerosity encoding follows Weber's law (e.g., Whalen et al., 1999). These findings are also in accordance with recent theories suggesting that visual numerosity is extracted independently of low-level features using a segmented visual input, at least for a moderate numerosity array (Adriano, Rinaldi, & Girelli, 2021; Adriano, Girelli, & Rinaldi, 2021; Anobile et al., 2014, 2016, 2017; Kirjakovski & Matsumoto, 2016).

Because we equalised the SF and luminance histogram of the stimuli across numerosities, one may wonder which type of visual information the visual system may exploit to segment the individual objects in the collection and compute the approximate numerosity. It should be noted that our main control of low-level stimuli features was specifically aimed to equalise only the

amplitude spectrum (e.g., SF, amplitude and orientations) and histogram luminance statistics across numerosities (and ratios), but the original *phase* of the stimuli was preserved (Railo et al., 2016; Joubert et al., 2009). According to the Fourier theory of spatial visual perception, any two-dimensional images can be decomposed into the sum of a set of sinusoidal gratings, as described by four basic parameters: orientation, SF, amplitude and phase (De Valois & De Valois, 1990). After performing the Fourier transformation on an image, we can thus obtain two key pieces of information: the *amplitude* spectrum (e.g., difference in luminance between the lightest and darkest parts of each constituent grating at a given SF and orientation) and the *phase* spectrum (e.g., the position relative to a reference point of each sinusoid at a given SF and orientation). Shortly, any image signal would be composed of its amplitude information and phase information (e.g., De Valois & De Valois, 1990). It has been suggested that the amplitude spectrum (or the squared value, the power spectrum) encodes image intensity or the global energy of the image (e.g., luminance, contrast, SF and orientation), whereas the phase spectrum encodes the spatial distribution of that energy and information about local image structures such as edges and contours, since edges require phase alignment across SF components (e.g., Arsenault, Yoonessi, & Baker, 2011; Gladilin & Eils, 2015; Morrone & Burr, 1988). In other words, phase information would encode the edges and contours of an image or something similar to Marr's concept of Primal Sketch (Marr, 1982). Accordingly, it has been suggested that the phase spectrum contains the most perceptually important image information (Piotrowski & Campbell, 1982). For instance, several studies using stimuli equated in the amplitude spectrum suggest that early object processing largely relies on phase information (Bieniek, Pernet, & Rousselet, 2012; Drewes, Wichmann, & Gegenfurtner, 2006; Joubert et al., 2009; Wichmann, Braun, & Gegenfurtner, 2006; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010); moreover, it has been suggested that early event-related potentials to faces and objects are mainly due to phase information rather than to the amplitude spectrum (e.g., Bieniek et al., 2012).

Therefore, independently of the role of phase information for object processing, here we solely maintain that power spectrum content *alone* cannot explain our results, but some other encoded visual information could likely be the source that participants used to perform the tasks (e.g., Wichmann et al., 2006). Indeed, we cannot advocate that the amplitude spectrum has no contribution to the task at all. As it has been shown for other non-numerical tasks (Joubert et al., 2009; Girard & Koenig-Robert, 2011), we found that SF averaging in numerical tasks may simply cause a drop in the accuracy and RTs compared to the original stimuli. In sum, the power spectrum

per se may only play a partial role in number processing, nevertheless it may improve the performance and precision of numerical representation (e.g., smaller CoV) when correlated/confounded with numerosity, as in the original images (as found in Experiment 2). This is also in line with previous studies investigating the role of other continuous cues, suggesting that numerical precision (e.g., CoV) is improved when visual cues are confounded with numerosity compared to when they are controlled/not confounded (Smets et al., 2015). Consequently, it is possible that when the power spectrum covaries or is confounded with numerosity (as in the original generated stimuli for our experiments), texture models alone may equally capture the performance. In addition, as suggested by some studies focusing on object and scene processing, when both amplitude and phase are available these two types of information are not used alone but likely interact (Drewes et al., 2006; Gaspar & Rousselet, 2009).

Several pieces of evidence suggest that texture models (Dakin et al., 2011; Durgin, 2008), or raw SF information alone, cannot always explain performance in numerical tasks. Further evidence in favour of segmentation accounts come from studies employing the so-called connectedness illusions (Franconeri et al., 2009; He et al., 2009; He et al., 2015). For example, it has been shown that connecting objects (e.g., dots) with task-irrelevant items led to a numerosity underestimation despite continuous features being kept almost constant across connectedness levels. The connected dot-pairs created by the irrelevant lines proportionally reduced the perceived numerosity because the visual system likely processes two connected dots as a single unified perceptual object, as suggested by the grouping principle of element connectedness (Palmer & Rock, 1994). Recently, Kirjakovski and Matsumoto (2016) replaced real lines with illusory contour lines to reduce visual confounds introduced by the physical lines (e.g., item visibility and further distractions) and they reported a numerosity underestimation proportional to the number of connections. In addition, Adriano and colleagues (Adriano, Rinaldi, & Girelli, 2021), using the same manipulation based on ICs, reported an underestimation effect even though the spatial profiles of the stimuli (e.g., SF) and other continuous features (e.g., density, convex hull, etc.) were carefully matched across connectedness levels, suggesting that numerosity processing may exploit segmented objects rather than mere raw texture-statistics (see also Adriano, Girelli, & Rinaldi, 2021). Furthermore, Anobile and colleagues (2017), in line with texture-density models, predicted that adding irrelevant physical lines to the stimulus would increase the energy in the high SFs (and the physical density) of the array, which in turn should make the array appear numerically larger than a numerically equal array of isolated dots. However, as they demonstrated, adding irrelevant lines that connected few dots

actually increased the energy in the high SFs (and the physical density) of the stimuli, but the manipulation led to a decrease rather than to an increase of both the perceived numerosity and the perceived density. These results clearly challenge the model of Dakin et al. (2011), but only over a moderate numerosity range (e.g., 15 items). Indeed, when the same connectedness manipulation was applied to a collection of 100 dots, the numerosity/density was overestimated following the predictions of Dakin's energy model, thus suggesting that texture-density mechanisms may be involved in highly cluttered arrays.

This range of studies, hence, support recent emergent proposals suggesting that collections of objects are not always processed by texture-density mechanisms, but different contextual-dependent visual mechanisms may coexist. For example, for numerically higher arrays (e.g., ≈ 200 items, texture-density), where items are too cluttered or crowded and object segmentation becomes harder, texture-density may be at play, whereas other systems are activated for very small numerosities (e.g., 1-4 items, or subitising) or for moderate (> 4, ANS) numerical arrays (Anobile et al., 2014, 2016; Kirjakovski & Matsumoto, 2016; Pomè, Anobile, Cicchini, Scabia, & Burr, 2019). Specifically, Anobile and colleagues (2014) reported that the Weber fraction was constant for moderate arrays and then decreased with the square root of numerosity for very dense arrays, thus suggesting a texture-density related mechanism for higher numerosities. Furthermore, a recent work (Anobile, Turi, Cicchini, & Burr, 2015) suggests that the switch between these systems can be regulated by crowding mechanisms depending on the visual eccentricity of the stimuli. This notion is also supported by another work (Valsecchi, Toscani, & Gegenfurtner, 2013) reporting that the numerosity of peripherally viewed dot arrays is reduced relative to the numerosity perceived in the fovea, specifically when arrays were quite dense and clustered (but see also Chakravarthi, & Bertamini, 2020 for an alternative explanation). To rule out confounding factors that covary with the stimulus eccentricity, these studies also manipulated SF by means of blurring the image. However, no strong effect of blurring was found over numerosity, suggesting that the reduced numerosity in periphery was specifically due to crowding mechanisms affecting dot individuation (see also Burgess & Barlow, 1983 who found an effect of blurring only for high levels of blur). By contrast, another study (e.g., Morgan et al., 2014; Experiment 3) independently manipulated blur level and numerosity in the same experiment and identified a similar threshold in the discrimination of blurring (e.g., which stimulus is more blurred) and numerosity (e.g., which stimulus contains more blobs) in the observers, a result that would suggest a rather strict connection between numerosity discrimination and other visual image properties. In line with the proposal of Dakin and colleagues

(2011), the work by Morgan and colleagues (2014) suggests that numerosity might be directly correlated with the energy in the high SFs. Indeed, as we increase the number of items in an image, we also add more contour. Hence, the amount of contour can be estimated from the combined output of 'edge detectors' that respond to local changes in luminance, meaning measuring the energy at high SFs in Fourier-optical terms. Therefore, the work of Anobile et al. (2014, 2015, see also Anobile et al., 2017) partially supports the central idea of the study of Morgan et al. (2014) but only over high numerical arrays, especially when dots are densely packed and subjected to crowding.

Although Burgess and Barlow (1983) measured estimates over a wide range of numerosities (e.g., up to 400 items) and showed that the Weber fraction decreased with numerosity, hence proposing a simple power variability model, more recent studies discovered a constant Weber fraction (or CoV) at least over less dense, moderate numerical arrays. For example, a recent study (Pomè, Anobile, Cicchini, & Burr, 2019) shows a pattern largely consistent with a constant CoV over the range 8 – 80 (see also Anobile et al., 2014, range 3 – 30) and another recent study found similar results (DeWind, Bonner, & Brannon, 2020) over the range 8 – 32. Similarly, Adriano et al. (2021) found that numerical acuity was constant in the range 9 – 15 in a numerical estimation task. We nevertheless pinpoint that this is a long-lasting question (i.e., the "Fechner-Weber-Stevens" debate) pertaining not only to the encoding of numerosity, but also to the exact shape of the psychophysical function relating to the intensity of every physical (continuous) magnitude with the internal subjective response (e.g., Krueger, 1989; Dehaene, 2003). Therefore, the fact that JND (and Weber's law) was not a central notion in Fechner's original derivations would provide a bridge toward modern psychophysics and facilitate the generalisation of his principles to the discrimination of multidimensional stimuli (see Dzhafarov & Colonius, 2011). Tackling this issue is beyond the scope of our work, although recently it has been suggested that strict adherence to Weber's law (e.g., JND increases linearly with numerosity) may depend on contextual factors varying from experiment to experiment, including also continuous sensory cues confounded with numerosity (for a review, see Testolin & McClelland, 2020). For instance, it has been shown that the effect of continuous sensory cues over performance can be modulated by stimulus duration, changes in instructions, task difficulty and task context (Leibovich, Henik, & Salti, 2015; Leibovich-Raveh, Stein, Henik, & Salti, 2018). Thus, three different systems may be flexibly recruited for numerosity processing, but perhaps contextual factors such as the number of items in the scene and other task constraints (e.g., stimulus exposure time) could activate one of these systems or promote different strategies based

on varying visual information, such as the discrete numerosity, or rather on low-level visual cues (e.g., Dietrich, Nuerk, Klein, Moeller, & Huber, 2019).

In sum, it is likely that global texture-models of numerosity processing (Dakin et al., 2011; Durgin, 2008) are well suited to explaining participants' performance under specific visual contexts such as cluttered scenes, ultra-rapid stimulus exposure duration, and with sets of objects in which the power spectrum naturally covaries or is confounded with numerosity. Nevertheless, as we showed in the current study, in other specific visual contexts (e.g., when the power spectrum is experimentally uncorrelated with the numerosity) segmentation models proposing a Weber-like encoding of numerosity may equally explain the performance (Burr & Ross, 2008a; Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012).

7.6. Conclusions

In sum, the present study demonstrates that numerosity processing is broadly independent from low-level visual features such as the power spectrum and luminance, at least for a moderate numerical range. We indeed overcame any possible methodological confounds by using stimuli carefully matched for SF amplitude spectrum and luminance distribution, and by independently varying the numerical ratio keeping other low-level features (e.g., density and convex hull, etc.) uncorrelated across numerosities. In two experiments, we discovered Weber-like numerical behaviour (ratio effect) although the power spectrum was completely uninformative of numerosity. The removal of power spectrum information merely caused a drop in the performance but not a total disruption of the ratio effect. Taken together, the current work strengthens the notion that, at least for a moderate numerosity range, the ANS does not exploit summary texture statistics or the global power spectrum *per se.*

Chapter 8

8.1. Spatial frequency equalization does not prevent spatial-numerical associations

Chapter adapted from: Adriano, A., Rinaldi, L., & Girelli, L. (2021). *Psychonomic Bulletin & Review*. In press.

8.2. Introduction

Recent works employing non-symbolic dot-arrays comparison tasks in adult participants reported the presence of SNAs compatible with those observed in infants and animals (Nemeh, Humberstone, Yates, & Reeve, 2018; Zhou, Shen, Li, Li, & Cui, 2016). For instance, Zhou and colleagues (2016) used a same/different matching task to investigate the spatial representation of non-symbolic numerosity, controlling also for low-level visual features such as size and density typically confounded with numerosity. Results showed that neither size nor density affected responses; yet, results yielded faster right-hand responses to large non-symbolic numerosities only (no difference was found for left-hand responses). Furthermore, Nemeh et al. (2018), using a target-to-reference comparison task, showed typical SNAs congruency effect (although arrays were not controlled for any low-level feature).

Yet, a very recent theoretical proposal, named as the *brain's asymmetric frequency tuning* (BAFT) hypothesis, suggested that SNAs would simply reflect laterality differences in the way the brain processes specific physical features in the actual numerical stimuli (Felisatti et al., 2020a, b). In particular, this hypothesis assumes that SNAs would emerge as the result of brain asymmetries relative to the processing of the raw spatial frequencies (SF) content naturally correlated with dot numerosity stimuli (or any other visual image), with the left or right brain hemisphere preferentially dedicated to process high or low SF bands, respectively. SFs are generally defined as the number of dark-light cycles per degree of visual angle (or per image). Low SFs (few cycles per degree) capture the global distribution of light and dark across the entire scene; high SFs (many cycles per degree) instead code local changes from light to dark that correspond to smaller elements (e.g., De Valois & De Valois, 1990). Since small numerical arrays of dots would contain fewer local changes from light

to dark (e.g., edges), they would be ideally represented by SF-defined contrast gratings with a few large strips per degree (e.g., low SF spectrum), while large arrays would contain more local dark/light variations and would be ideally represented by grating with many thin strips per degree (e.g., high SF spectrum). Hence, non-symbolic SNAs would be the result of the lateralized SF processing in each hemisphere. For example, in the case of adult participants, a visually presented small array of dots would engage more the right hemisphere, hence inducing a left bias and speeding up the manual response with the left hand (Felisatti et al., 2020a, b). Accordingly, since in the study of Nemeh and colleagues (2018) dot size was constant, SFs information were strictly correlated with dot numerosity, hence leaving open the possibility that their results could be explained by the BAFT account.

Crucially, the BAFT account makes very specific testable experimental predictions since it assumes that, overall, SNAs are driven by the physical content of the stimuli such as their SF power spectrum (e.g., De Valois & De Valois, 1990). This means that when this physical information is not informative about numerosity depicted in the stimuli, SNAs should not be observed (e.g., Wichmann, Drewes, Rosas, & Gegenfurtner, 2010). Here, to directly test this hypothesis, we run two experiments on adult participants: in Experiment 1, they were required to perform a typical sequential non-symbolic comparison task (Nemeh et al., 2018) with classic arrays of dots (i.e., original stimuli), but controlled for 5 main low-level features. In striking contrast, in Experiment 2 we removed SF information as a cue for numerosity by equalizing the full spectrum across all numerical dot-arrays (see also Adriano, Girelli, & Rinaldi, 2021a, b). According to the BAFT hypothesis, if SNAs originate from the brain asymmetrical tuning in processing raw SF content, SNAs effect should be found with the original stimuli, but not with the SF-equalized ones. Alternatively, if SF information does not play a key role, we should expect spatial-numerical compatibility effects in both experiments, as well as a typical ratio effect (e.g., Whalen et al., 1999).

8.3. Experiment 1: Comparison Task with Original stimuli

In Experiment 1 we tested whether non-symbolic dots arrays may trigger SNA using a classic dots comparison task. Participants were presented with a stimulus (reference) followed by a second one (test), and they had to decide whether the latter was numerically smaller or larger than the reference (for a similar task, see Nemeh et al., 2018). Therefore, unlike previous studies that did not fully control low-level features correlated with numerosity (Nemeh et al., 2018; Zhou et al., 2016), here we used dot arrays controlled for 5 main visual features: convex hull, total surface, density, item size and total circumference (Gebuis & Reynvoet, 2011). We manipulated the numerical ratio between reference and test stimuli and the mapping of response keys: in one condition, the mapping was congruent (e.g., "smaller" was associated with a left response and "larger" with a right response) whereas in the other condition it was incongruent. We predicted that if SNAs emerge from spatial coding of non-symbolic numerical information, we should observe a typical ratio effect and a typical congruency mapping effect (e.g., slower RTs for incongruent mapping).

8.3.1. Materials and methods

8.3.1.1. Participants

We performed an a priori power analysis with G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to determine our needed sample size. Because our study was inspired by the work of Nemeh et al. (2018), who found a large effect size (e.g., $\eta^2_p = .19$) for the typical SNA congruency effect (e.g., hand x magnitude interaction), we assumed a standard large effect size ($\eta^2_p = .14$) for our main variable of interest (e.g., congruency mapping). The calculation established that, in order to obtain a large effect size ($\eta^2_p = .14$) with an 80% of power for the main effect of the mapping (congruent vs incongruent), in a one-way repeated measures ANOVA (2 levels of measurements; alpha = .05), a minimum sample of 51 participants was required.

A sample of 52[9] undergraduate students from the University of Milano-Bicocca were recruited (41 females, 44 right-handed). The mean age was 22.53 years (*SD* = 3.69). Due to Covid

[9] A total of 62 participants were originally tested. However, data were inspected before the analysis to check that participants correctly understood the task. Consequently, 10 participants were discarded from the final sample because presented an overall accuracy outside the ±1.5 interquartile range of the distribution.

restrictions, participants performed the study online through the Pavlovia/PsychoPy platform (www.pavlovia.org). All participants had normal or corrected-to-normal vision and were unaware of the purpose of the experiment. Each subject signed an online informed consent document before the experiment began and the study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Local Ethical Committee (protocol N° RM-2020-230).

8.3.1.2. Stimuli

Original stimuli (Figure 8.1) were generated off-line with the script from Gebuis and Reynvoet (2011), as in the previous chapter (see *Chapter* 7).

Figure 8.1: Example of original stimuli used in Experiment 1 as generated with the method of Gebuis and Reynvoet (2011).

8.3.1.3. Procedure

Instructions and experimental stimuli were projected by means of an online PsychoPy routine (Peirce, 2007). The experimental task was a number comparison between two sequentially presented arrays of dots (e.g., to determine whether the test stimulus is numerically larger or smaller than the fixed reference). The experiment was preceded by a brief training period composed of 8 trials allowing participants to familiarise with the task. In the training phase, we presented only the condition with the lowest (i.e., easiest) ratio 0.66. Each trial started with a blank screen for 500 ms (RGB = 0, 0, 0), before a grey fixation cross (Font: Times; Size: 16 pixels; RGB = 127, 127, 127) was presented for an additional 500 ms and followed by a further blank screen (500 ms). Next, test stimulus was displayed within a black window (RGB = 0, 0, 0) on the screen centre for 300 ms;

afterwards, a blank screen (500 ms) was presented until the onset of test stimuli, which stay on the screen until response (Figure 8.2).

After the training phase, two experimental blocks composed of 96 randomised trials were presented, for a total of 192 experimental trials. In one block (congruent mapping), participants were instructed to press the left key ("A" key) with their left index finger if they judged the test numerosity to be smaller than reference or to press the right key ("L" key) with their right index finger if they judged the test numerosity to be larger than reference. In the other block, the mapping of the keys was reversed (incongruent mapping). The order of blocks was counterbalanced across subjects. Within a block, each of the 6 comparison pairs (test vs reference) were repeated 16 times, resulting in 96 total trials per block (16 trials × 6 comparison pairs).

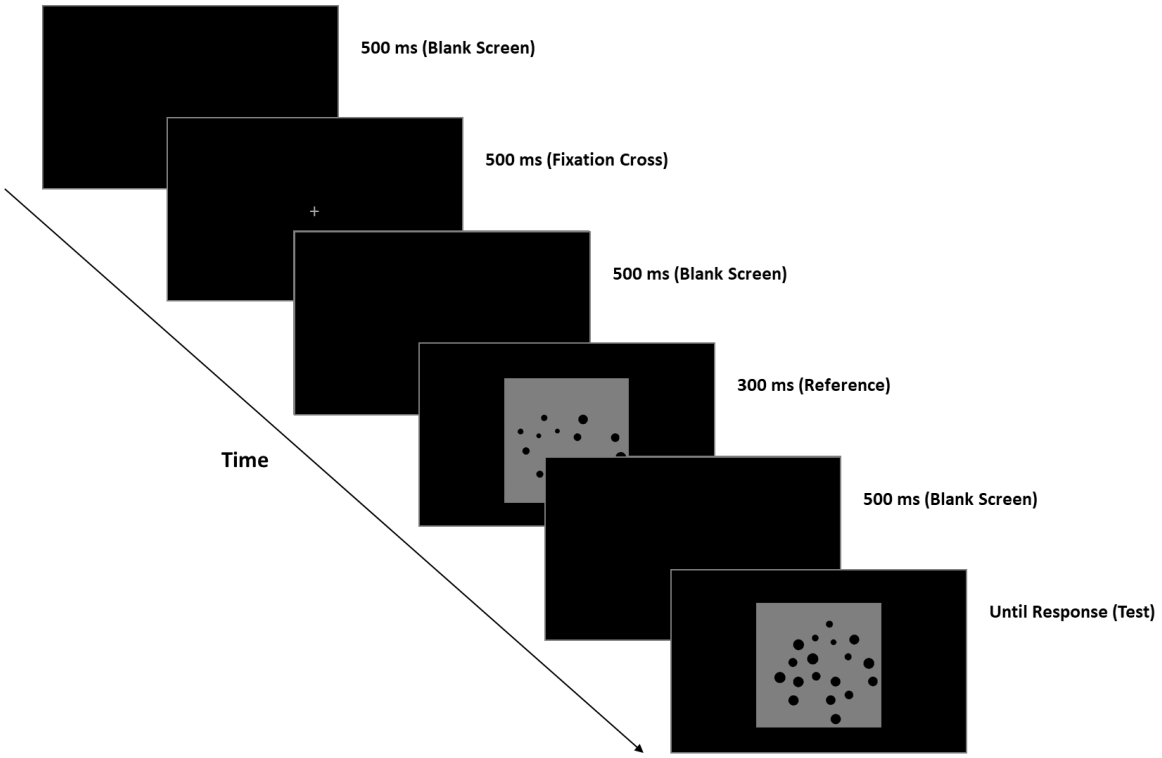


Figure 8.2: The number comparison task. The participant had to decide whether the test stimulus was numerically larger or smaller than reference stimulus. A total of 192 trials (96 trials × 2 blocks) were displayed.

8.3.2. Results and Discussion of Experiment 1

Two separated 2 x 3 repeated-measures ANOVAs were performed with response mapping (congruent vs incongruent) and numerical ratio (0.66, 0.75, 0.8) as within-subject factors and with RTs or accuracy (percentage of correct responses) as dependent variable, respectively[10]. The analyses of accuracy data only showed a main effect of numerical ratio, $F(2, 102) = 187.11$, $\varepsilon = .89$, $p < .001$, $\eta^2_p = .78$, suggesting that discrimination was harder for higher ratios (Figure 8.3A). Post-hoc test (Bonferroni correction), revealed a significant difference between the ratio 0.66 and 0.75, $t(102) = 4.774$, $p < .001$, $d = .66$, the ratio 0.66 and the ratio 0.80, $t(102) = 18.622$, $p < .001$, $d = 2.58$, and the ratio 0.75 and the ratio 0.80, $t(102) = 13.84$, $p < .001$, $d = 1.9$. On the contrary, no main effect of mapping, $F(1, 51) = .009$, $p = .92$, $\eta^2_p = .001$, nor interaction was found, $F(2, 102) = .57$, $p = .52$, $\eta^2_p = .003$.

The RTs analysis on correct responses (data were log-transformed and trials outside ±1.5 times the interquartile range of distribution were eliminated, for a total of 3.14% datapoints removed) showed a main effect of numerical ratio, $F(2, 102) = 41.2$, $\varepsilon = .88$, $p < .001$, $\eta^2_p = .44$, with numerosity discrimination becoming slower for harder ratios (Figure 8.3B). Post-hoc comparisons (Bonferroni correction) revealed a significant difference between the ratio 0.66 and 0.75, $t(102) = -4.772$, $p < .001$, $d = .66$, the ratio 0.66 and the ratio 0.80, $t(102) = -9.075$, $p < .001$, $d = 1.25$, and the ratio 0.75 and the ratio 0.80, $t(102) = -4.303$, $p < .001$, $d = .59$. Crucially, we also found a significant main effect of mapping, $F(1, 51) = 11.03$, $p = .002$, $\eta^2_p = .178$, with faster responses for congruent compared to incongruent mapping. No interaction was found, $F(2, 102) = .98$, $\varepsilon = .88$, $p = .36$, $\eta^2_p = .019$.

To corroborate these results, we also run a further analysis following Fias, Brysbaert, Geypens and d'Ydewalle (1996). Reaction times with left hand responses were subtracted from those with the right hand and were fitted with a linear regression as a function of each numerosity tested (Fias et al., 1996). Results showed that numerosity explained a significant proportion of variance in the differential RTs (e.g., right-left), $R^2 = .90$, $F(1, 4) = 35.1$, $p < .001$. As expected, we found a significant negative regression coefficient, $\beta = -.02$, $t(4) = -5.92$, $p < .001$, suggesting a linear mapping: for numerosities smaller than reference, RTs were faster for the left hand and vice versa

[10] The Greenhouse-Geisser epsilon (ε) correction for violation of sphericity was applied when needed and original *F*, *df* and corrected *p*-values were reported.

for numerosities larger than reference (Figure 8.3C; see also Supplementary Materials for the individual analyses, which are in line with the results reported here).

As a further independent index of numerical acuity, we also calculated the Coefficient of Variation (CoV) for each participant and mapping condition, as an index of the Weber fraction (e.g., Halberda & Odic, 2014). Gaussian cumulative distribution functions were fitted to the data (e.g., proportion of test stimuli correctly judged as more numerous than the reference, as a function of numerosity in test stimuli) and parameters were estimated with a parametric approach based on the maximum likelihood method, using *Quickpsy* package for *R* (Linares & López-Moliner, 2016). Psychometric curves were fitted considering the typical lapse in performance (e.g., missing a trial, finger-errors) by allowing the value of the guess rate (γ) and lapse rate (λ) parameters to vary in the default range of 0 – 0.05 (Wichmann & Hill, 2001). The CoV was computed as the ratio between the standard deviation (SD) and the mean (e.g., point of subjective equality; PSE) of the psychometric functions (e.g., Helbig & Ernst, 2007). In line with the overall results of accuracy, we did not find a significant difference in the CoV between congruent and incongruent conditions, $t(51) = -.324$, $p = .74$, $d = 0.045$ (Figure 8.3D). Frequentist analyses were accompanied by Bayesian statistics that confirmed these results (see Supplementary materials).

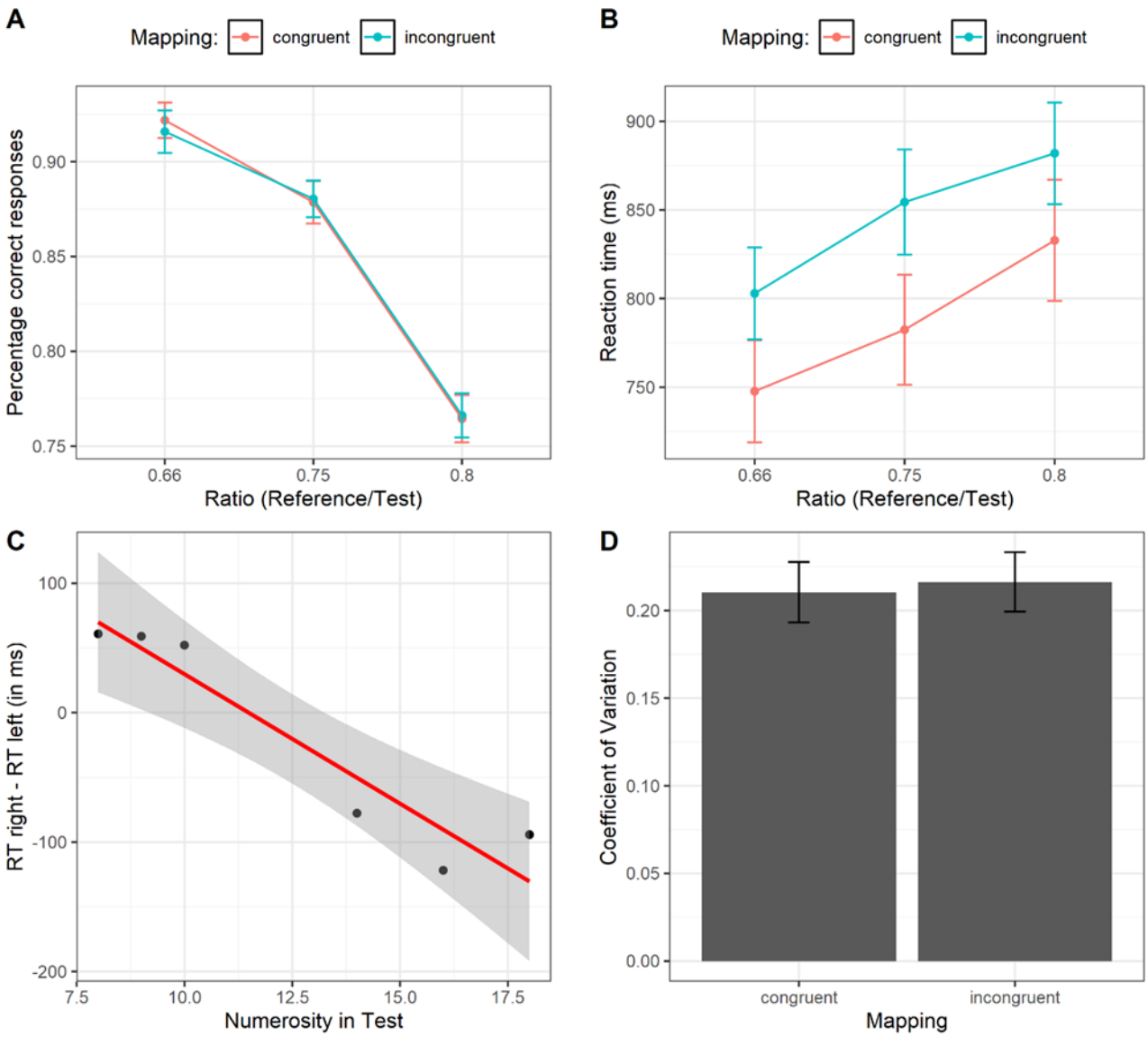


Figure 8.3: A) Percentage of correct responses as a function of the absolute ratio and the mapping condition. B) Reaction times as a function the absolute ratio and the mapping condition. Bars represent ±1 SEM. C) RT difference between responses with the right and left hands as a function of the numerosity in test stimuli. Shaded regions represent the 95% CI of the regression line. D) Coefficient of Variation for each mapping condition. Bars represent ±1 SEM.

8.4. Experiment 2: Comparison Task with Spatial Frequency equalized stimuli

In Experiment 2, we specifically tested whether the SNAs merely emerge from the raw SFs content of the stimuli as predicted by the BAFT hypothesis. Participants were tested in a comparison task as in the Experiment 1, but in this case, stimuli were equalized for SF content. According to the BAFT hypothesis, when this information is removed as cue for numerosity (e.g., it was equalized across all stimuli) we should expect no ratio effect and, crucially, no SNA (Felisatti et al., 2020a, b). That is, since all the reference and test stimuli have the same power spectrum, participants cannot use this cue during the task to classify test stimuli as "larger" or "smaller" than the reference: hence, the performance should be merely at chance, showing no difference across mapping conditions (e.g., RTs for left-hand responses should be equal for smaller *and* larger numerosities, with such pattern extending to RTs for the right-hand responses). On the other hand, if SNA does not depend on SF content alone, the compatibility effect should still emerge (e.g., RTs for left-hand responses should be faster for smaller numerosities and slower for larger numerosities, while the opposite pattern should be observed for the right-hand responses).

8.4.1. Materials and methods

8.4.1.1. Participants

A new sample of 52[11] undergraduate students from the University of Milano-Bicocca were recruited (42 females, 50 right-handed). The mean age was 22.34 years (*SD* = 2.74). All the participants performed the study online through the Pavlovia/PsychoPy platform.

8.4.1.2. Stimuli & Procedure

Stimuli and procedure were identical to those of Experiment 1. The only difference is that original stimuli where post-processed using a Matlab script, following a similar methodology of Adriano et al. (2021a, b; *see Chapter 6 and 7*). Hence, each stimulus had a similar SFs amplitude spectrums and luminance profiles across numerosities and ratios.

[11] A total of 60 subjects were originally tested. Following data inspection, 8 participants were discarded from the final sample because presented an overall accuracy outside the ±1.5 interquartile range of the distribution.

8.4.2. Results and Discussion of Experiment 2

Data were analyzed as in Experiment 1. We found a main significant effect of the ratio, $F(2, 102) = 207.92$, $\varepsilon = .88$, $p < .001$, $\eta^2_p = .80$, Figure 8.4A, and a main effect of mapping over accuracy, $F(1, 51) = 5.9$, $p = .019$, $\eta^2_p = .104$, but no interaction, $F(2, 102) = .238$, $\varepsilon = .83$, $p = .74$, $\eta^2_p = .005$. Post-hoc (Bonferroni correction), revealed a significant difference between the ratio 0.66 and 0.75, $t(102) = 6.631$, $p < .001$, $d = .92$, the ratio 0.66 and the ratio 0.80, $t(102) = 20.01$, $p < .001$, $d = 2.7$, and the ratio 0.75 and the ratio 0.80, $t(102) = 13.38$, $p < .001$, $d = 1.85$.

Analysis of RTs (4.65 % of data were discarded) showed a main effect of ratio, $F(2, 102) = 43.22$, $p < .001$, $\eta^2_p = .45$. Post-hoc comparisons (Bonferroni correction), revealed a significant difference between the ratio 0.66 and 0.75, $t(102) = -3.01$, $p = .010$, $d = .41$, the ratio 0.66 and the ratio 0.80, $t(102) = -9.12$, $p < .001$, $d = 1.26$, and the ratio 0.75 and the ratio 0.80, $t(102) = -6.11$, $p < .001$, $d = .84$. Crucially, a significant main effect of mapping was also found, $F(1, 51) = 4.57$, $p = .037$, $\eta^2_p = .082$, Figure 8.4B, with faster RTs for the congruent mapping as compared to the incongruent one[12]. No interaction was found, $F(2, 102) = .72$, $p = .48$, $\eta^2_p = .014$.

Finally, we replicated the regression analysis with the RTs difference between right- and left-hand responses as dependent variable, and numerosity as predictor. Again, we found that numerosity explained a significant proportion of variance in RTs, $R^2 = .93$, $F(1, 4) = 52.99$, $p = .001$, with a significant decreasing regression coefficient, $\beta = -.01$, $t(4) = -7.28$, $p = .001$, suggesting a linear mapping (Figure 8.4C; see also Supplementary Materials for the individual analyses). As a further metric of numerical precision, we calculated the CoV (an index of Weber fraction) for each mapping condition. We found that in the congruent condition participants presented also a slightly better precision (smaller CoV) compared to the incongruent condition, $t(51) = -2.25$, $p = .029$, $d = .31$, Figure 8.4D, which is in line also with higher accuracy found for the congruent condition compared to the incongruent condition. Frequentist analyses were accompanied by Bayesian statistics that confirmed these results (see Supplementary materials).

[12] We also run an overall ANOVA across the two experiments, hence including the type of stimulus (i.e., original vs. equalized) as a between-subjects factor. Results on RTs showed only a significant main effect of numerical ratio and, critically, of mapping. Furthermore, the congruency effect was not statistically different across the two Experiments (original vs. equalized). See the Supplementary Materials for further information.

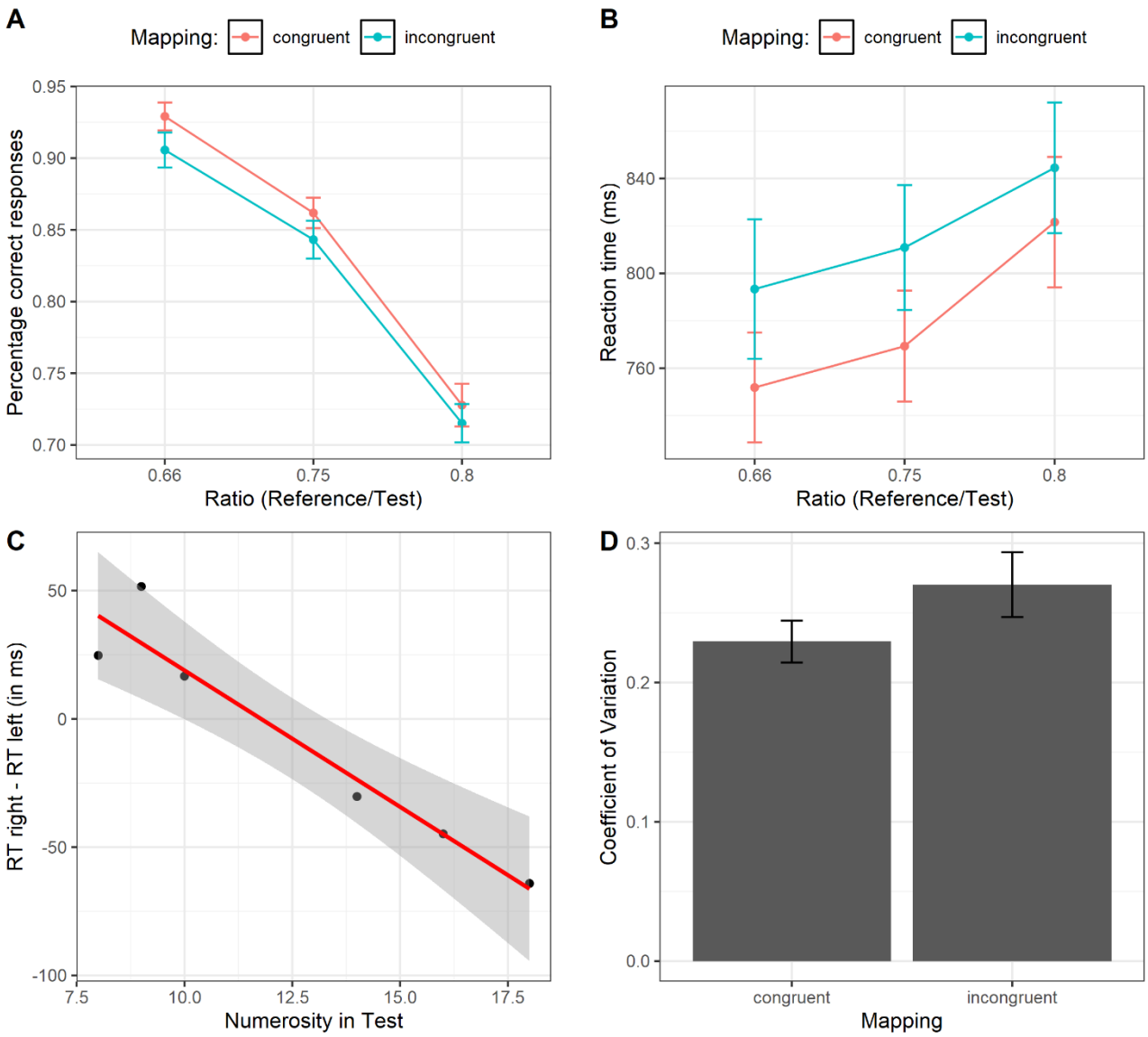


Figure 8.4: A) Percentage of correct responses as a function of the absolute ratio and the mapping condition. B) Reaction times as a function the absolute ratio and the mapping condition. Bars represent ±1 SEM. C) RT difference between right- and left-hand responses as a function of the numerosity in test stimuli. Shaded regions represent the 95% CI of the regression line. D) Coefficient of Variation for each mapping condition. Bars represent ±1 SEM.

8.5. General Discussion

In this study, we directly tested the *brain's asymmetric frequency tuning* hypothesis (Felisatti et al., 2020a, b), probing whether non-symbolic spatial-numerical associations originate from a mere spatial frequency coding of the raw visual input. Results from two experiments revealed the presence of a spatial-numerical association with non-symbolic numerosity information regardless of whether spatial frequencies were equalized or not. Indeed, in Experiment 2, we completely ruled out the role of SF, since both congruency-mapping and ratio effect were replicated also when the full power spectrum was equalized across all stimuli. According to this hypothesis, any smaller/larger numerosity would be naturally associated with lower/higher SF content, and this would determine the observed spatial mapping (as the left/right hemisphere preferentially processes higher/lower SF information). In other terms, BAFT theory provides a strong neurological explanation of the spatial-numerical association effect, but it represents a general visual mechanism for the processing of any visual image, including also numerosity. Accordingly, the presence of SNAs across development in preliterate children (de Hevia et al., 2014; Bulf et al., 2016), in human adults (Nemeh et al., 2018; Zhou et al., 2016), as well as in animals (Rugani et al., 2015; 2020), particularly in studies using dot stimuli, is assumed to be the mere result of the lateralization of neural structures devoted to process SFs information composing any visual image. For example, in human adults the right hemisphere should be tuned for low spatial frequencies and, accordingly, stimuli with low SF information would be preferentially processed by this part of the brain, that also control for the contralateral left hand, inducing a left-bias (and viceversa for the left hemisphere). Thus, according to BAFT account, since the right hemisphere is tuned for low SF, the speed of response with the (contralateral) left hand should be fastened depending on the low SF content of a visual image presented at the center of the visual field. This should explain why visually presented small numerical arrays of dots (e.g., low SF information) should be judged faster with the left hand, compared to larger arrays of dots (e.g., high SF information). However, in Experiment 2 all the numerical test stimuli were matched for the whole power spectrum, which means that physical information reaching the right hemisphere was constant across all numerical arrays and mapping conditions. Hence, with power spectrum equalized stimuli under the BAFT account hypothesis, the reaction times for left hand responses should have been similar for stimuli depicting larger or smaller numerosities, since both visual sets contained the same (low) raw SFs information amount. Yet, and contrarily to BAFT predictions, in the Experiment 2 we specifically found that latencies for

left hand responses were still modulated by the *numerical content* depicted in the arrays (encoded in the original phase of the stimuli), rather than by the raw power spectrum. Indeed, we found a typical congruency mapping, which means that RTs for left hand were faster for smaller numerosities and slower for larger numerosities (and viceversa for the right-hand). While several models have been proposed to explain the SNAs (for a review see Van Dijck, Ginsburg, Girelli, & Gevers, 2015), the BAFT model is perhaps the most reductionist among them, since does not assume any particular cognitive (e.g., numerical mental line representation) or attentive factor behind the whole process, and is rather rooted on a strict number of computational factors: the lateralization of brain structures processing different SFs ranges and, crucially, the physical information contained in the raw visual input stimulus. In a neuro-computational metaphor, any algorithm trained to classify numerical stimuli (e.g., as larger or smaller) extracting *only* their raw power spectrum, would be "tricked" if tested with our SF-spectrum equalized stimuli (Wichmann et al., 2010) and would simply fail to do the task. If our brain implements a similar processing mode, real observers should have failed completely the comparison task and no SNA and ratio effect should have been found in the Experiment 2. Therefore, our results challenge the role of a mere SF processing since this information was cancelled out as cue for numerosity. In that respect, cancelling out the power spectrum information and still observing a SNA, help us to reject the BAFT account among the several proposed models in the literature, leaving open the possibility that spatial-numerical link with non-symbolic arrays of dots might emerge according to other cognitive factors (e.g., extraction of approximate numerical information and mapping across a spatially oriented mental representation). These findings are in line with prior psychophysical studies showing that SF power spectrum *alone* cannot explain the typical behavioral effects observed in non-symbolic numerical processing, such as the ratio dependence and scalar variability (e.g., Whalen et al., 1999), at least for moderate arrays of numerosity (Adriano et al., 2021a, b; Anobile, Cicchini, & Burr, 2014; Anobile et al., 2017). Similarly, numerosity discrimination seems preserved also in animals when SFs are controlled in the numerical stimuli (Potrich, Zanon, & Vallortigara, 2021). Indeed, the fact that power spectrum could be *correlated* with dot numerosity does not necessary imply that the visual system takes into account *only* this feature to process numerosity magnitude (Wichmann et al., 2010). Accordingly, here we clearly found that when we control for the (natural) correlation between global power spectrum and numerosity, the performance is not dramatically impaired (i.e., typical ratio effect is observed) and a clear SNAs can therefore be found.

8.6. Conclusions

In sum, these data indicate that SNAs emerging with non-symbolic numerosities, at least in human adults, cannot be traced back to brain asymmetries relative to the processing of SFs power spectrum, leaving open the possibility that a combination of *attentional* and *numerical* factors could be at the origin of such effect and that cultural experiences may play a role in shaping this effect (de Hevia, 2021).

Our work is in line with studies reporting spatial-numerical associations in which numerosity (i.e., in the form of non-symbolic arrays) were task-relevant (Nemeh et al., 2018; Zhou et al., 2016), whereas there is little consensus when numerosity is not required to be explicitly estimated (e.g., numerical Posner-like task). Indeed, while some studies found that observing a relatively large numerical dot-array fasten saccades toward the subsequent target presented in the right space and vice versa for small numerical arrays (e.g., Bulf et al., 2016), others failed to find similar SNAs-like effects with both task-irrelevant symbolic and non-symbolic stimuli (e.g., Fattorini, Pinto, Rotondaro, & Doricchi, 2015; Cleland, Corsico, White, & Bull, 2020; see also Colling et al., 2020). The methodology we employed here can be easily integrated also with these paradigms to understand whether discrepancies, at least among studies using only dot arrays, might be due to SFs confounds, hence paving the way for further work aimed at understanding the origin of the spatial-numerical associations.

Chapter 9

9.1. Final Discussion

The Number sense theory (e.g., Dehaene & Changeux, 1993) has represented for several years the most influential model within the Numerical Cognition field. Several lines of research have established that these "innate" numerical abilities are rooted in specific neural structures mainly localized in parietal lobes (although also early visual cortex and frontal cortex might be involved in the process) allowing live beings to extract abstract numerical information from the sensory input (e.g., "approximate" numerosity). However, this theory has been recently challenged by several findings that seem to contradict the Number sense theory. Indeed, if numerosity is an abstract property, why non-numerical visual factors, such as item proximity, item density, or convex-hull still apparently influence the performance?

Hence, the question about how numerosity is computed by the visual system has captivated cognitive scientists over the last decades, becoming a core theoretical dispute concerning the type of stimulus features and related neurocognitive mechanisms that may play a key role in number representation. In particular, the last years have seen a rise in popularity of theories maintaining that numerosity representation would be indirectly extracted through surrogate low-level features correlated or confounded with the numerosity in the set (e.g., Gebuis et al., 2016; Leibovich et al., 2017). This has led to new accounts challenging previous theoretical hypotheses, according to which the ANS would be based on a primitive visual segmentation and individuation algorithm and, hence, on discrete mechanisms (Burr & Ross, 2008; Dehaene & Changeux, 1993).

In this thesis, taking advantage of classic visual illusions and modern computational methods, we tried to challenge alternative models of numerosity perception showing that the Number Sense theory is still the most parsimonious theoretical model that account for number processing. Indeed, we demonstrated that perception of numerosity is a flexible process depending on several contextual factors that can modulate how the visual scene is segmented, with numerosity computation that likely starts already at an early stage (e.g., V2) of the visual input processing (e.g., before assigning a "numerical value" to the organized percept).

In particular, in **Chapter 3** we experimentally demonstrated that numerosity perception is regulated by the number of connected pairs in the scene. Indeed, even when continuous cues were

equalized across ICs conditions, perceived numerosity still varied with the number of illusorily connected “dumbbell pairs”. This strongly suggests that numerosity might be based on a global-to-local perception of the scene in which perceptual organization may play a fundamental role. In this sense, visual numerosity can be also seen as a second-order information (Clarke & Beck, 2021), emerging after perceptual organization and item segmentation have been accomplished (as demonstrated by the “connectedness” effect). Furthermore, in **Chapter 4**, we showed that underestimation effects triggered by the illusory-connecting lines are not due to brightness confounds. Indeed, we show that a boundary grouping mechanism, insensitive to inducer contrast, might be the actual origin of the observed underestimation effect. This, hence, confirms the results of the previous chapter suggesting that the effect is specifically due to the increased grouping strength and not due to uncontrolled confounds (e.g., perceived brightness). In **Chapter 5** we also investigated the effect of *perceived size* of the arrays (e.g., convex-hull) when set numerosity is embedded in Ponzo contexts, while we independently varied the number of ICs. We showed that perceived numerosity scaled according to the contextual cues of the Ponzo illusion, as suggested by several previous studies. However, contrary to the conclusions of previous studies in the literature (according to which the building blocks of numerosity are mere continuous features, e.g., see Gebuis et al., 2016; Leibovich et al., 2017), we clearly showed that numerosity perception was still modulated by the number of connections, even when perceived visual cues were constant. This may suggest that discrete segmentation mechanisms persist but, perhaps, they could be affected by spatial information (perhaps at an early input level), explaining in turn previous findings in the literature on the role of continuous variables. In **Chapter 6** we also investigated the correlation between numerosity and low/high SFs in the arrays, in both comparison and estimation tasks. We show that breaking this correlation by equalizing the power spectrum across all numerical stimuli and ICs conditions did not prevent discrete numerosity perception since the performance was still affected by the number of illusorily connected dot-pairs. In **Chapter 7**, similarly, using classic dot arrays controlled for several individual visual features (e.g., convex-hull, density, etc.), we also showed that the ratio-based performance, a hallmark of the Weber-like encoding of numerosity, does not originate from mere SF processing code *alone*. Indeed, we clearly found that observers’ performance depended on the ratio between numerosities, even when power spectrum was constant across stimuli pairs. Accordingly, we suggest that phase information may partially have a role on numerosity processing. Finally in **Chapter 8** we showed that non-symbolic spatial numerical associations, a behavioral signature of the (innate) internal mental number line, are not based on

SF processing and/or brain asymmetries in their processing. In accordance with the original SNAs reported with symbolic digits, we found that response to non-symbolic small dot numerosities were faster with the left hand and to large dot numerosities with the right hand, even when SFs were equalized across conditions.

Taken together, these studies strongly suggest that the Number Sense theory is still valid in its essence, even though some refinement perhaps is needed to the original model. Indeed, we stress here that the role of *perceptual organization* is overlooked in the original (computational) theory and in more recent "alternative accounts". We suggest that numerosity is neither based on the number of individual local closed surfaces in the scene, nor on a mere sum of low-level visual features. Rather, several visual parameters affect how numerosity can be perceived according to perceptual (Gestalt) organization laws. More than one hundred years of Gestalt psychology have established that our visual system is indeed extremely adapted to segment the visual scene into figure/ground relations (e.g., Wagemans et al., 2012).

It is worth noting that numerosity perception and, in particular, visual numerosity illusions have been largely investigated at the beginning of the Gestalt school, but for several years the field has been quiescent (for a review see Luccio, 2016). Only recently, we assisted to a continuously increasing number of studies focused on perceptual aspects of numerosity processing using visual illusions and grouping rules. It follows that future (new) models of numerosity processing should account for the effects of several visual factors that may affect how individual local items are segmented, such as item-proximity (e.g., Allik & Raidvee, 2021; Chakravarthi & Bertamini, 2020), homogeneity (Redden & Hoch, 2009; DeWind et al., 2020), physical connectedness (Franconeri et al., 2009; He et al., 2009; He et al., 2015), illusory connectedness (Kirjakovski & Matsumoto, 2016) and several other spatial factors (Dormal et al., 2018; Pecunioso et al., 2020; Picon et al., 2019). In short, we suggest that the Number sense model should be extended integrating also the knowledge gained in visual perception and Gestalt psychology. For example, it is not unlikely that item-proximity may simply affect how individual items are segmented by affecting the grouping strength. Hence this could explain the underestimation found when items are physically closer. In sum, item proximity may affect numerical process in the same way that connectedness (and other gestalt rules such as similarity, etc) affects how the items are grouped into coherent perceptual objects (e.g., Chakravarthi & Bertamini, 2020).

Furthermore, for reaching a better understanding of the role of low-level features in numerical processing, future research should investigate real-world visual contexts and not simply

“laboratory” manipulations. Indeed, despite the fact that when the objects are spatially scattered without any overlap among them, continuous variables can be actually correlated with numerosity (thus representing a possible indirect proxy for numerosity), it should be noted that in real-world perception the physical information that define an object (e.g., edges) cannot be always completely visible in the visual scene, due to several visual and biological constraints. In fact, in the 3D-world, it is very likely to perceive and interact with objects that lack some “portion” either because they are partially occluded by other objects or because they have an unclear figure-ground relationship, as when they have a similar colour of the background. Furthermore, the retinal projections are interrupted by blind spots and retinal veins. Because of both these visual constraints in 3-D world perception and the intrinsic structure of the eyes, the number and extent of visible “fragmented” surfaces impressed on the retina, cannot directly convey the number of objects in the external world (cf. Riesen, Ruda, & Mingolla, 2014). That is, the continuous variables comprising the stimuli could not be a useful proxy-information to generate a “reliable” numerical magnitude estimation in every situation, as simply captured on the retinal projection of the objects (Franconeri et al., 2009). Recently, it has also been shown that dyscalculic children have a larger crowding effect compared to controls (Castaldi et al., 2020). Since a strict link may exist between crowding and grouping/segmentation (Francis, Manassi, & Herzog, 2017) in visual processing, it would be not surprising if further perceptual weakness may unravel and explain the origin of some deficits in dyscalculic subjects, such as impaired ANS representation or worse Weber fraction (precision) in non-symbolic numerical tasks.

References

Aagten-Murphy, D., Attucci, C., Daniel, N., Klaric, E., Burr, D., & Pellicano, E. (2015). Numerical estimation in children with autism. *Autism Research*, *8*(6), 668–681.

Adriano, A., Girelli, L., & Rinaldi, L. (2021). Non-symbolic numerosity encoding escapes spatial frequency equalization. *Psychological Research*, 1–14.

Adriano, A., Girelli, L., & Rinaldi, L. (2021). The ratio effect in visual numerosity comparisons is preserved despite spatial frequency equalisation. *Vision Research*, *183*, 41–52.

Adriano, A., Rinaldi, L., & Girelli, L. (2021). Visual illusions as a tool to hijack numerical perception: Disentangling nonsymbolic number from its continuous visual properties. *Journal of Experimental Psychology: Human Perception and Performance, 47*(3), 423–441.

Agrillo, C., & Bisazza, A. (2018). Understanding the origin of number sense: A review of fish studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740), 20160511.

Agrillo, C., Dadda, M., & Bisazza, A. (2007). Quantity discrimination in female mosquitofish. *Animal Cognition*, *10*(1), 63-70.

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Animal Cognition*, *11*(3), 495-503.

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2009). Use of number by fish. *PloS one*, *4*(3), e4786

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*(2), 281-287.

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PloS one*, *7*(2), e31923.

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2015). Ratio dependence in small number discrimination is affected by the experimental procedure. *Frontiers in Psychology*, *6*, 1649.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.

Allik, J., & Raidvee, A. (2021). Proximity model of perceived numerosity. *Attention, Perception, & Psychophysics*, 1-10.

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*(4), 303–314.

Allik, J., Tuulmets, T., & Vos, P. G. (1991). Size invariance in visual number discrimination. *Psychological research*, *53*(4), 290-295.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245–266.

Anobile, G., Arrighi, R., Togoli, I., & Burr, D. C. (2016). A shared numerical representation for action and perception. *eLife*, *5*, e16161.

Anobile, G., Burr, D. C., Iaia, M., Marinelli, C. V., Angelelli, P., & Turi, M. (2018). Independent adaptation mechanisms for numerosity and size perception provide evidence against a common sense of magnitude. *Scientific Reports*, *8*(1), 1–12.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science*, *25*(1), 265–270.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception*, *45*(1-2), 5–31.

Anobile, G., Cicchini, G. M., Gasperini, F., & Burr, D. C. (2018). Typical numerosity adaptation despite selectively impaired number acuity in dyscalculia. *Neuropsychologia*, *120*, 43–49.

Anobile, G., Cicchini, G. M., Pomè, A., & Burr, D. C. (2017). Connecting visual objects reduces perceived numerosity and density for sparse but not dense patterns. *Journal of Numerical Cognition*, *3*(2), 133–146.

Anobile, G., Turi, M., Cicchini, G. M., & Burr, D. C. (2015). Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *Journal of Vision*, *15*(5), 1–12.

Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: an event-related functional magnetic resonance imaging study. *Journal of cognitive neuroscience*, *18*(11), 1820-1828.

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child development*, 695-701.

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20141791.

Arsenault, E., Yoonessi, A., & Baker, C. (2011). Higher order texture statistics impair contrast boundary segmentation. *Journal of Vision*, *11*(10), 1–15.

Baumann, R., van der Zwan, R., & Peterhans, E. (1997). Figure-ground segregation at contours: A neural mechanism in the visual cortex of the alert monkey. *European Journal of Neuroscience*, *9*(6), 1290–1303.

Benson-Amram, S., Heinen, V. K., Dryer, S. L., & Holekamp, K. E. (2011). Numerical assessment and individual call discrimination by wild spotted hyaenas, Crocuta crocuta. *Animal Behaviour*, *82*(4), 743–752.

Beran, M. J., Evans, T. A., Leighty, K. A., Harris, E. H., & Rice, D. (2008). Summation and quantity judgments of sequentially presented sets by capuchin monkeys (Cebus apella). *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *70*(2), 191-194.

Bertamini, M., Zito, M., Scott-Samuel, N. E., & Hulleman, J. (2016). Spatial clustering and its effect on perceived clustering, numerosity, and dispersion. *Attention, Perception, & Psychophysics*, *78*(5), 1460–1471.

Bevan, W., Maier, R. A., & Helson, H. (1963). The influence of context upon the estimation of number. *The American journal of psychology*, *76*(3), 464-469.

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: Evidence from parametric, test-retest, single-subject analyses. *Journal of Vision*, *12*(13), 1–24.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, *282*(5389), 746–749.

Brannon, E. M., Abbott, S., & Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition*, *93*(2), B59–B68.

Brus, J., Heng, J. A., & Polanía, R. (2019). Weber's law: A mechanistic foundation after two centuries. *Trends in Cognitive Sciences*, *23*(11), 906–908.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*(6), 1131–1136.

Bulf, H., de Hevia, M. D., & Macchi Cassia, V. (2016). Small on the left, large on the right: Numbers orient visual attention onto space in preverbal infants. *Developmental Science*, *19*(3), 394–401.

Burgess, A., & Barlow, H. B. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research*, *23*(8), 811–820.

Burr, D. C., & Ross, J. (2008a). A visual sense of number. *Current Biology*, *18*(6), 425–428.

Burr, D. C., & Ross, J. (2008b). Response: Visual number. *Current Biology*, *18*(18), R857–R858.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological science*, *17*(5), 401–406.

Cantlon, J. F., & Brannon, E. M. (2007). How much does number matter to a monkey (Macaca mulatta)?. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(1), 32–41.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS biology*, *4*(5), e125.

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition*, *128*(3), 331–352.

Castaldi, E., Turi, M., Gassama, S., Piazza, M., & Eger, E. (2020). Excessive visual crowding effects in developmental dyscalculia. *Journal of Vision*, *20*(8), 1–20.

Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Sciences*, *103*(12), 4693–4698.

Chakravarthi, R., & Bertamini, M. (2020). Clustering leads to underestimation of numerosity, but crowding is not the cause. *Cognition*, *198*, 104195.

Chen, S., Glasauer, S., Müller, H. J., & Conci, M. (2018). Surface filling-in and contour interpolation contribute independently to Kanizsa figure formation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(9), 1399–1413.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications*, *7*(1), 1–7.

Cipora, K., Soltanlou, M., Reips, U. D., & Nuerk, H. C. (2019). The SNARC and MARC effects measured online: Large-scale assessment methods in flexible cognitive effects. *Behavior Research Methods*, *51*(4), 1676–1692.

Clarke, S., & Beck, J. (2021). The Number Sense Represents (Rational) Numbers. *Behavioral and Brain Sciences*, 1-57.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological science*, *10*(5), 408-411.

Cleland, A. A., Corsico, K., White, K., & Bull, R. (2020). Non-symbolic numerosities do not automatically activate spatial–numerical associations: Evidence from the SNARC effect. *Quarterly Journal of Experimental Psychology*, *73*(2), 295-308.

Colling, L. J., Szűcs, D., De Marco, D., Cipora, K., Ulrich, R., Nuerk, H. C., ... & Langton, S. R. (2020). A multilab registered replication of the attentional SNARC effect. *Advances in Methods and Practices in Psychological Science*, *3*(2), 143-162.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic bulletin & review*, *8*(4), 698-707.

Coren, S., & Ward, L. M. (1979). Levels of processing in visual illusions: The combination and interaction of distortion-producing mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 324–335.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences*, *108*(49), 19552–19557.

Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature*, *371*(6500), 791–793.

de Hevia, M. D. (2021). How the human mind grounds numerical quantities on space. *Child Development Perspectives*, *15*(1), 44–50.

de Hevia, M. D., Girelli, L. & Macchi Cassia, V. (2012). Minds without language represent number through space: Origins of the mental number line. *Frontiers in Psychology*, *3*, 466.

de Hevia, M. D., Girelli, L., Addabbo, M., & Macchi Cassia, V. (2014). Human infants' preference for left-to-right oriented increasing numerical sequences. *PloS one*, *9*(5), e96412.

de Hevia, M. D., Vallar, G., & Girelli, L. (2008). Visualizing numbers in the mind's eye: The role of visuo-spatial processes in numerical abilities. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1361–1372.

de Hevia, M. D., Veggiotti, L., Streri, A., & Bonn, C. D. (2017). At birth, humans associate “few” with left and “many” with right. *Current Biology*, *27*(24), 3879–3884.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. New York: Oxford University Press.

Decarli, G., Paris, E., Tencati, C., Nardelli, C., Vescovi, M., Surian, L., & Piazza, M. (2020). Impaired large numerosity estimation and intact subitizing in developmental dyscalculia. *Plos one*, *15*(12), e0244578.

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147.

Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*(2), 219-250.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive neuropsychology*, *20*(3-6), 487–506.

Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology, 14*(4), 613–636.

DeWind, N. K. (2019). The number sense is an emergent property of a deep convolutional neural network trained for object recognition. *bioRxiv*, 609347.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247-265.

DeWind, N. K., Bonner, M. F., & Brannon, E. M. (2020). Similarly oriented objects appear more numerous. *Journal of Vision*, *20*(4), 1–11.

DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex*, *114*, 76–89.

Dietrich, J. F., Huber, S., & Nuerk, H. C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS)–a research review. *Frontiers in Psychology*, *6*, 295.

Dietrich, J. F., Nuerk, H. C., Klein, E., Moeller, K., & Huber, S. (2019). Set size influences the relationship between ANS acuity and math performance: A result of different strategies? *Psychological Research*, *83*(3), 590–612.

Ditz, H. M., & Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. *Proceedings of the National Academy of Sciences*, *112*(25), 7827–7832.

Dormal, V., Larigaldie, N., Lefèvre, N., Pesenti, M., & Andres, M. (2018). Effect of perceived length on numerosity estimation: Evidence from the Müller-Lyer illusion. *Quarterly Journal of Experimental Psychology*, *71*(10), 2142–2151.

Dresp, B., & Grossberg, S. (2016). Neural computation of surface border ownership and relative surface depth from ambiguous contrast inputs. *Frontiers in Psychology*, *7*, 1102.

Dresp, B., Salvano-Pardieu, V., & Bonnet, C. (1996). Illusory form with inducers of opposite contrast polarity: Evidence for multistage integration. *Perception & Psychophysics*, *58*(1), 111–124.

Drewes, J., Wichmann, F. A., & Gegenfurtner, K. R. (2006). Classification of natural scenes: Critical features revisited. *Journal of Vision*, *6*(6), 561a.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, *80*(1), 61–95.

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology*, *18*(18), R855–R856.

Dzhafarov, E. N., & Colonius, H. (2011). The Fechnerian idea. *American Journal of Psychology*, *124*(2), 127–140.

Ebersbach, M., Luwel, K., & Verschaffel, L. (2014). Further evidence for a spatial-numerical association in children before formal schooling. *Experimental Psychology*, *61*, 323–329

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, *247*(4943), 721–723.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.

Evers, K., de-Wit, L., Van der Hallen, R., Haesen, B., Steyaert, J., Noens, I., & Wagemans, J. (2014). Brief report: Reduced grouping interference in children with ASD: Evidence from a multiple object tracking task. *Journal of Autism and Developmental Disorders*, *44*(7), 1779–1787.

Evers, K., Van der Hallen, R., Noens, I., & Wagemans, J. (2018). Perceptual organization in individuals with autism spectrum disorder. *Child Development Perspectives*, *12*(3), 177–182.

Fattorini, E., Pinto, M., Rotondaro, F., & Doricchi, F. (2015). Perceiving numbers does not cause automatic shifts of spatial attention. *Cortex*, *73*, 298-316.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Felisatti, A., Laubrock, J., Shaki, S., & Fischer, M. H. (2020a). A biological foundation for spatial–numerical associations: The brain's asymmetric frequency tuning. *Annals of the New York Academy of Sciences*, *1477*(1), 44–53.

Felisatti, A., Laubrock, J., Shaki, S., & Fischer, M. H. (2020b). Commentary: A mental number line in human newborns. *Frontiers in Human Neuroscience*, *14*, 99.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2*(1), 95–110.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Research*, *33*(2), 173–193.

Fornaciai, M., & Park, J. (2018). Early numerosity encoding in visual cortex is not sufficient for the representation of numerical magnitude. *Journal of Cognitive Neuroscience*, *30*(12), 1788–1802.

Fornaciai, M., & Park, J. (2021). Disentangling feedforward versus feedback processing in numerosity representation. *Cortex*, *135*, 255–267.

Fornaciai, M., Brannon, E. M., Woldorff, M. G., & Park, J. (2017). Numerosity processing in early visual cortex. *Neuroimage*, *157*, 429–438.

Francis, G., Manassi, M., & Herzog, M. H. (2017). Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychological Review*, *124(4),* 483–504.

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*(1), 1–13.

Frith, C. D., & Frith, U. (1972). The solitaire illusion: An illusion of numerosity. *Attention, Perception, & Psychophysics*, *11*(6), 409–410.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65.

Galton, F. (1880a). Visualised numerals. *Nature*, *21*, 252–256.

Galton, F. (1880b). Visualised numerals. *Nature*, *21*, 494–495.

Gaspar, C. M., & Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Research*, *49*(24), 3001–3012.

Geary, D. C. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychological bulletin*, *114*(2), 345.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*(4), 981–986.

Gebuis, T., & Reynvoet, B. (2012a). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642–648.

Gebuis, T., & Reynvoet, B. (2012b). The role of visual information in numerosity estimation. *PloS one*, *7*(5), e37426.

Gebuis, T., & Reynvoet, B. (2012c). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology*, *49*(11), 1649–1659.

Gebuis, T., Cohen Kadosh, R. C., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica*, *171*, 17–35.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: local diagnostic value. *Archives of Neurology & Psychiatry*, *44*(2), 398–408.

Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy. *The American Journal of Psychology*, 267-273.

Ginsburg, N., & Goldstein, S. R. (1987). Measurement of visual cluster. *The American Journal of Psychology*, *100*(2), 193–203.

Girard, P., & Koenig-Robert, R. (2011). Ultra-rapid categorization of Fourier-spectrum equalized natural images: Macaques and humans perform similarly. *PLoS One*, *6*(2), e16453.

Gladilin, E., & Eils, R. (2015). On the role of spatial phase and phase correlation in vision, illusion, and cognition. *Frontiers in Computational Neuroscience*, *9*, 45.

Grossberg, S. (2014). How visual illusions illuminate complementary brain processes: Illusory depth from brightness and apparent motion of illusory contours. *Frontiers in Human Neuroscience*, *8*, 854.

Grossberg, S., & Mingolla, E. (1987). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. In *The adaptive brain II* (pp. 143-210). Elsevier.

Grossberg, S., & Mingolla, E. (1987). The role of illusory contours in visual segmentation. In S. Petry & G. E. Meyer (Eds.), *The perception of illusory contours* (pp. 116–125). New York: Springer.

Grossberg, S., & Raizada, R. D. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, *40*(10-12), 1413–1432.

Grossberg, S., & Repin, D. V. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. *Neural Networks*, *16*(8), 1107–1140.

Halberda, J., & Odic, D. (2014). The precision and internal confidence of our approximate number thoughts. In D. C. Geary, D. Berch, & K. Koepke (Eds.), *Evolutionary origins and early development of number processing* (pp. 305–333). San Diego: Academic Press.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*(6150), 1123–1126.

He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review*, *16*(3), 509–517.

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences*, *112*(41), E5647–E5655.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, *179*(4), 595–606.

Henschen SL (1919): On language, music and calculation mechanisms and their localisation in the cerebrum. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 52: 273– 298.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435-448.

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, *103*(51), 19599–19604.

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in Human Neuroscience*, *5*, 150.

Hyde, D. C., & Spelke, E. S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of cognitive neuroscience*, *21*(6), 1039-1053.

Im, H. Y., Zhong, S. H., & Halberda, J. (2016). Grouping by proximity and the visual impression of approximate number in random dot arrays. *Vision Research*, *126*, 291–307.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 281–282

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision*, *9*(1), 1–16.

Karmel, B. Z. (1969). The effect of age, complexity, and amount of contour on pattern preferences in human infants. *Journal of Experimental Child Psychology*, *7*(2), 339-354.

Katzin, N., Katzin, D., Rosén, A., Henik, A., & Salti, M. (2020). Putting the world in mind: The case of mental representation of quantity. *Cognition*, *195*, 104088.

Kiefer, M., & Dehaene, S. (1997). The time course of parietal activation in single-digit multiplication: Evidence from event-related potentials. *Mathematical cognition*, *3*(1), 1-30.

Kirjakovski, A., & Matsumoto, E. (2016). Numerosity underestimation in sets with illusory contours. *Vision Research*, *122*, 34–42.

Kluth, T., & Zetzsche, C. (2016). Numerosity as a topological invariant. *Journal of Vision*, *16*(3), 1–39.

Koesling, H., Carbone, E., Pomplun, M., Sichelschmidt, L., & Ritter, H. (2004, May). When more seems less-non-spatial clustering in numerosity estimation. In *Early Cognitive Vision Workshop ECOVISION* (Vol. 28).

Kogo, N., Drożdżewska, A., Zaenen, P., Alp, N., & Wagemans, J. (2014). Depth perception of illusory surfaces. *Vision Research*, *96*, 53–64.

Kogo, N., Strecha, C., Van Gool, L., & Wagemans, J. (2010). Surface construction by a 2-D differentiation–integration process: A neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. *Psychological Review*, *117*(2), 406–439.

Kramer, P., Di Bono, M. G., & Zorzi, M. (2011). Numerosity estimation in visual stimuli in the absence of luminance-based cues. *PloS one*, *6*(2), e17378.

Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, *11*(1), 5-9.

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences*, *12*(2), 251–267.

Laeng, B., & Endestad, T. (2012). Bright illusions reduce the eye's pupil. *Proceedings of the National Academy of Sciences*, *109*(6), 2162–2167.

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, 375.

Leibovich, T., & Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *The Quarterly Journal of Experimental Psychology*, *67*(5), 899–917.

Leibovich, T., Henik, A., & Salti, M. (2015). Numerosity processing is context driven even in the subitizing range: An fMRI study. *Neuropsychologia*, *77*, 137–147.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, *40,* e164.

Leibovich-Raveh, T., Stein, I., Henik, A., & Salti, M. (2018). Number and continuous magnitude processing depends on task goals and numerosity ratio. *Journal of Cognition*, *1*(1), 19.

Linares, D., & López-Moliner, J. (2016). Quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, *8*(1), 122–131.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological science*, *14*(5), 396-401.

Luccio, R. (2016). The Illusions of Numerosity. In A. G. Shapiro, & D. Todorovic (Eds.), *The Oxford compendium of visual illusions* (pp. 696 – 699), Oxford University Press.

Luria, A. R., 1966, *Higher Cortical Functions in Man.* New York: Basic Books.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*(1), 1–22.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman and Company.

Matthews, N., & Welch, L. (1997). The effect of inducer polarity and contrast on the perception of illusory figures. *Perception*, *26*(11), 1431–1443.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child development*, *82*(4), 1224-1237.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334.

Meck, W. H., & Church, R. M. (1984). Simultaneous temporal processing. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*(1), 1–29.

Miletto Petrazzini, M. E., & Agrillo, C. (2016). Turning to the larger shoal: Are there individual differences in small-and large-quantity discrimination of guppies?. *Ethology Ecology & Evolution*, *28*(2), 211-220.

Morgan, M. J., Raphael, S., Tibber, M. S., & Dakin, S. C. (2014). A texture-processing model of the 'visual sense of number'. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1790), 20141137.

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *235*(1280), 221–245.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519-1520.

Nasr, K., Viswanathan, P., & Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Science Advances*, *5*(5), eaav7903.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

Nemeh, F., Humberstone, J., Yates, M. J., & Reeve, R. A. (2018). Non-symbolic magnitudes are represented spatially: Evidence from a non-symbolic SNARC task. *PloS one*, *13*(8), e0203019.

Ni, C. F. (1934). The effect of combining some geometrical optical illusions. *The Journal of General Psychology*, *10*(2), 472–476.

Nieder, A. (2002). Seeing more than meets the eye: Processing of illusory contours in animals. *Journal of Comparative Physiology A*, *188*(4), 249–260.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382.

Nieder, A. (2021). Neuroethology of number sense across the animal kingdom. *Journal of Experimental Biology*, *224*(6), 1–15.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149-157.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences*, *101*(19), 7457–7462.

Nys, J., & Content, A. (2012). Judgement of discrete and continuous quantity in adults: Number counts! *The Quarterly Journal of Experimental Psychology*, *65*(4), 675–690.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255–265.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*(1), 29–55.

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2015). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex*, *26*(2), 748–763.

Parrish, A. E., Agrillo, C., Perdue, B. M., & Beran, M. J. (2016). The elusive illusion: Do children (*Homo sapiens*) and capuchin monkeys (*Cebus apella*) see the Solitaire illusion? *Journal of Experimental Child Psychology*, *142*, 83–95.

Pecunioso, A., Petrazzini, M. E. M., & Agrillo, C. (2020). Anisotropy of perceived numerosity: Evidence for a horizontal–vertical numerosity illusion. *Acta Psychologica*, *205*, 103053.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13.

Perdue, B. M., Talbot, C. F., Stone, A. M., & Beran, M. J. (2012). Putting the elephant back in the herd: Elephant relative quantity judgments match those of other species. *Animal Cognition*, *15*(5), 955–961.

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *Journal of Neuroscience*, *9*(5), 1749–1763.

Peterhans, E., & von der Heydt, R. (1991). Subjective contours: Bridging the gap between psychophysics and physiology. *Trends in Neurosciences*, *14*(3), 112–119.

Piaget, J. (1952). Play, dreams and imitation in childhood.

Piantadosi, S. T., & Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological Science*, *28*(4), 462–469.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33-41.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555.

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes?. *Neuroimage*, *15*(2), 435-446.

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain research*, *1106*(1), 177-188.

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, *24*(6), 1037–1043.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293-305.

Picon, E., Dramkin, D., & Odic, D. (2019). Visual illusions help reveal the primitives of number perception. *Journal of Experimental Psychology: General, 148*(10), 1675–1687.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage*, *14*(5), 1013–1026.

Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, *11*(3), 337–346.

Pomè, A., Anobile, G., Cicchini, G. M., Scabia, A., & Burr, D. C. (2019). Higher attentional costs for numerosity estimation at high densities. *Attention, Perception, & Psychophysics*, *81*(8), 2604–2611.

Pomè, A., Caponi, C., & Burr, D. C. (2021). Grouping-induced numerosity biases vary with Autistic-like personality traits. *Journal of Autism and Developmental Disorders*, 1–8.

Pomè, A., Thompson, D., Burr, D. C., & Halberda, J. (2020). Location and object-based attention enhance number estimation. *Attention, Perception, & Psychophysics*, *83*, 7–17.

Ponzo, M. (1928). Urteilstäuschungen über Mengen. *Archiv für die gesamte Psychologie*.

Poom, L., Lindskog, M., Winman, A., & Van den Berg, R. (2019). Grouping effects in numerosity perception under prolonged viewing conditions. *PloS one*, *14*(2), e0207502.

Porter, K. B., Mazza, V., Garofalo, A., & Caramazza, A. (2016). Visual object individuation occurs over object wholes, parts, and even holes. *Attention, Perception, & Psychophysics*, *78*(4), 1145–1162.

Portley, M., & Durgin, F. H. (2019). The second number-estimation elbow: Are visual numbers greater than 20 evaluated differently? *Attention, Perception, & Psychophysics, 81*(5)*,* 1512–1521.

Potrich, D., Zanon, M., & Vallortigara, G. (2021). Archerfish number discrimination. *bioRxiv*.

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology*, *17*(24), R1042-R1043.

Railo, H., Karhu, V. M., Mast, J., Pesonen, H., & Koivisto, M. (2016). Rapid and accurate processing of multiple objects in briefly presented scenes. *Journal of Vision*, *16*(3), 1–11.

Ramachandran, V. S. (1987). Visual perception of surfaces: A biological theory. In S. Petry & G. E. Meyer (Eds.), *The perception of illusory contours* (pp. 93–108). New York: Springer.

Redden, J. P., & Hoch, S. J. (2009). The presence of variety reduces perceived quantity. *Journal of Consumer Research*, *36*(3), 406-417.

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Attention, Perception, & Psychophysics*, *57*(7), 971–976.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1–3), 17–42.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*(6), 607–614.

Riesen, G., Ruda, H., & Mingolla, E. (2014). Rapidly estimating numerosity independent of size-related distance or occlusion. *Journal of Vision*, *14*(10), 879–879.

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics*, *73*(8), 2542–2572.

Ross, J., & Burr, D. (2012). Number, texture and crowding. *Trends in Cognitive Sciences*, *16*(4), 196–197.

RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Rugani, R., Vallortigara, G., Priftis, K., & Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Science*, *347*(6221), 534–536.

Rugani, R., Vallortigara, G., Priftis, K., & Regolin, L. (2020). Numerical magnitude, rather than individual bias, explains spatial numerical association in newborn chicks. *Elife*, *9*, e54662.

Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive development*, *10*(2), 253-269.

Smets, K., Sasanguie, D., Szücs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, *27*(3), 310–325.

Song, C., Schwarzkopf, D. S., & Rees, G. (2011). Interocular induction of illusory size perception. *BMC Neuroscience*, *12*(1), 27.

Sovrano, V. A., & Bisazza, A. (2008). Recognition of partly occluded objects by fish. *Animal Cognition*, *11*(1), 161–166.

Sovrano, V. A., & Bisazza, A. (2009). Perception of subjective contours in fish. *Perception*, *38*(4), 579–590.

Spehar, B. (2000). Degraded illusory contour formation with non-uniform inducers in Kanizsa configurations: the role of contrast polarity. *Vision Research*, *40*(19), 2653–2659.

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*(4473), 1033-1035.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*(2), 97-127.

Stoianov, I., & Zorzi, M. (2012). Emergence of a 'visual number sense' in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194–196.

Testolin, A., & McClelland, J. L. (2020). Do estimates of numerosity really adhere to Weber's law? A reexamination of two case studies. *Psychonomic Bulletin & Review*, 1–11.

The jamovi project (2019). *Jamovi* (Version 1.1.5) [Computer Software]. Retrieved from https://www.jamovi.org.

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), R11–R14.

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision*, *12*(6), 1–19.

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(36), 14513-14518.

Turi, M., Burr, D. C., Igliozzi, R., Aagten-Murphy, D., Muratori, F., & Pellicano, E. (2015). Children with autism spectrum disorder show reduced adaptation to number. *Proceedings of the National Academy of Sciences*, *112*(25), 78687872.

Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision*, *15*(4), 1–14.

Valsecchi, M., Toscani, M., & Gegenfurtner, K. R. (2013). Perceived numerosity is reduced in peripheral vision. *Journal of Vision*, *13*(13), 1–16.

Van Dijck, J. P., Ginsburg, V., Girelli, L., & Gevers, W. (2015). Linking numbers to space. *The Oxford handbook of numerical cognition*, 89–105.

van Lier, R., & Gerbino, W. (2015). Perceptual completions. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization* (pp. 294–320). New York: Oxford University Press.

Van Rinsveld, A., Guillaume, M., Kohler, P. J., Schiltz, C., Gevers, W., & Content, A. (2020). The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG. *Proceedings of the National Academy of Sciences*, *117*(11), 5726–5732.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*(4654), 1260–1262.

Vos, P. G., Van Oeffelen, M. P., Tibosch, H. J., & Allik, J. (1988). Interactions between area and numerosity. *Psychological Research*, *50*(3), 148-154.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in cognitive sciences*, *7*(11), 483–488.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*(2), 130–137.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.

Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research*, *46*(8-9), 1520–1529.

Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision*, *10*(4), 1–27.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.

Wilson, M. L., Britton, N. F., & Franks, N. R. (2002). Chimpanzees and the mathematics of battle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1496), 1107–1112.

Wurm, M. F., Porter, K. B., & Caramazza, A. (2019). Individuation of parts of a single object and multiple distinct objects relies on a common neural mechanism in inferior intraparietal sulcus. *Cortex*, *121*, 1–15.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749-750.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11.

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*(1), 88–101.

Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Attention, Perception, & Psychophysics*, *68*(5), 815–828.

Xu, Y. (2008). Representing connected and disconnected shapes in human inferior intraparietal sulcus. *Neuroimage*, *40*(4), 1849–1856.

Zavagno, D., Tommasi, L., & Laeng, B. (2017). The eye pupil's response to static and dynamic illusions of luminosity and darkness. *i-Perception*, *8*(4), 2041669517717754.

Zhang, J., Wu, Z., Wu, J., Mou, Y., & Yue, Z. (2020). The Effects of Auditory Numerosity and Magnitude on Visual Numerosity Representation: An ERP Study. *Journal of Numerical Cognition*, *6*(2), 164–185.

Zhou, X., Shen, C., Li, L., Li, D., & Cui, J. (2016). Mental numerosity line in the human's approximate number system. *Experimental Psychology*, *63,* 169–179.

Zimmermann, E., & Fink, G. R. (2016). Numerosity perception after size adaptation. *Scientific Reports*, *6*(1), 1–7.

Zorzi, M., &Butterworth, B. (1999). A computational model of number comparison. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 778–783). Mahwah, NJ: Erlbaum.Wilkey, E. D., & Ansari, D. (2019). Challenging the neurobiological link between number sense and symbolic numerical abilities. *Ann. NY Acad. Sci*, *40*, 1-23.

Supplementary Materials

Chapter 3

Supplementary Results: Experiment 1

In addition to the analyses reported in the main manuscript, we also ran two separate Bayesian repeated-measures ANOVAs respectively on the PSE and CoV, with the number of ICs as independent variable. For the PSE, results showed a BF_{10} = 32.77 for the effect of the number of ICs, suggesting a strong evidence in favour of the alternative hypothesis than of the null hypothesis (Table S3.1). For the CoV, we found a BF_{10} = .174 for the effect of the number of ICs, which suggests that data were 5.74 times more likely to occur under the null hypothesis (e.g., constant CoV across conditions) than under the alternative hypothesis (Table S3.2). Hence, the Bayesian analyses further corroborate the findings reported in the main manuscript.

In addition, we also explored whether the CoV (i.e., an index of discrimination precision) was stable at the within-individual level in the different ICs conditions, by calculating the correlation between individual CoV for each ICs condition. Results showed that all Pearson's correlation coefficients for each pair of variables (0 ICs-2 ICs; 0 ICs-4 ICs; 2 ICs-4 ICs) were significant (all $p < .05$), with a positive and strong relationship among the different illusory conditions (Figure S3.2 and Table S3.3). This means that subjects had a similar numerical discrimination precision across the three tested illusory conditions.

Table S3.1. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.030	0.031	1.000	
ICs	0.500	0.970	32.778	32.778	0.756

Note. All models include subject

Table S3.2. Bayesian ANOVA on the CoV.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.852	5.740	1.000	
ICs	0.500	0.148	0.174	0.174	1.051

Note. All models include subject

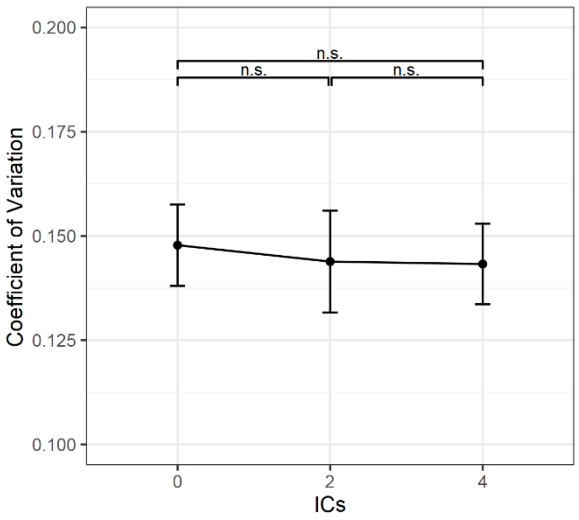


Figure S3.1: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

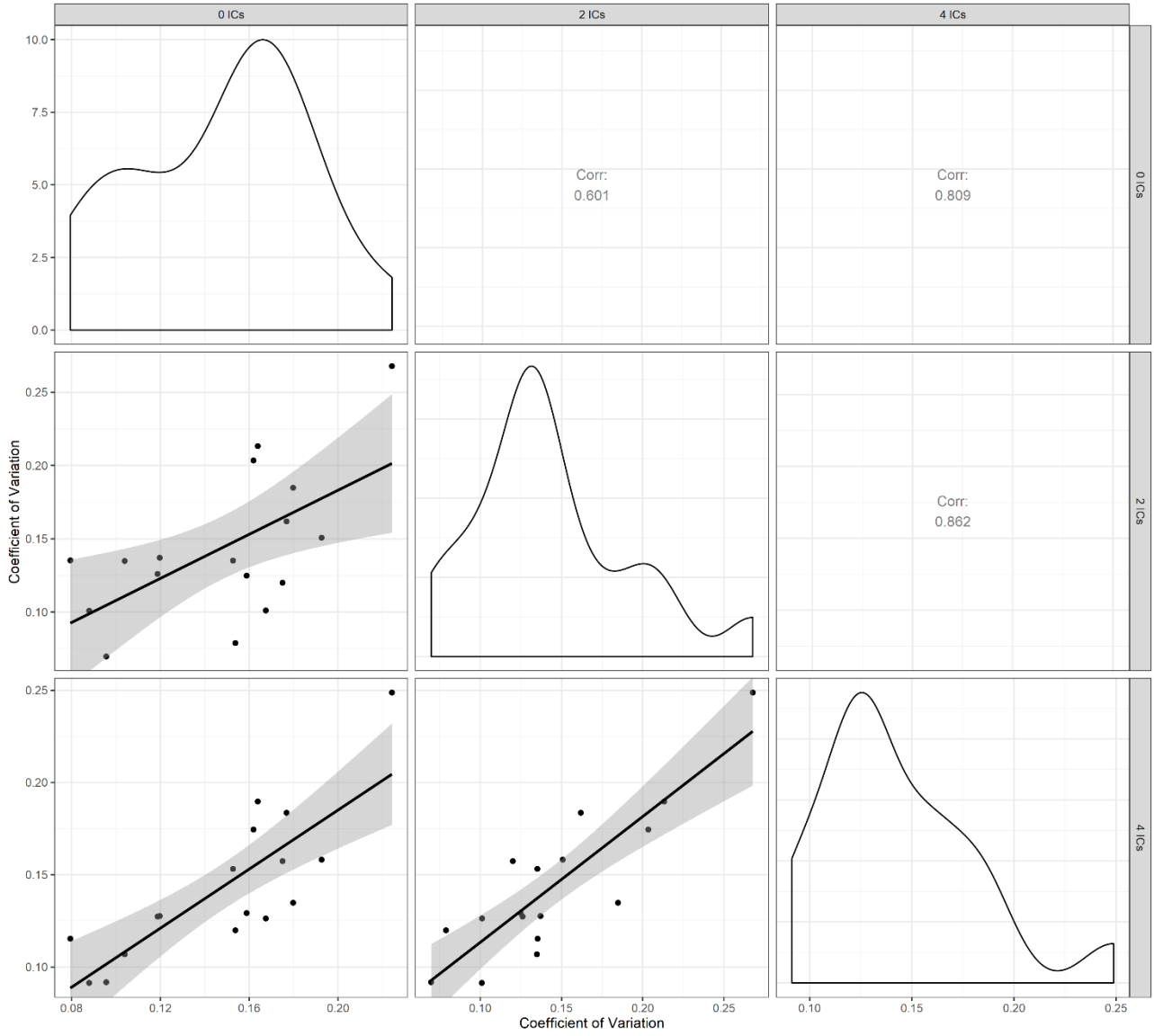


Figure S3.2: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Table S3.3. CoV correlation values among conditions.

Pearson Correlations

		0 ICs	2 ICs	4 ICs
0 ICs	Pearson's r	—		
	p-value	—		
2 ICs	Pearson's r	0.601 *	—	
	p-value	0.011	—	
4 ICs	Pearson's r	0.809 ***	0.862 ***	—
	p-value	< .001	< .001	—

* p < .05, ** p < .01, *** p < .001

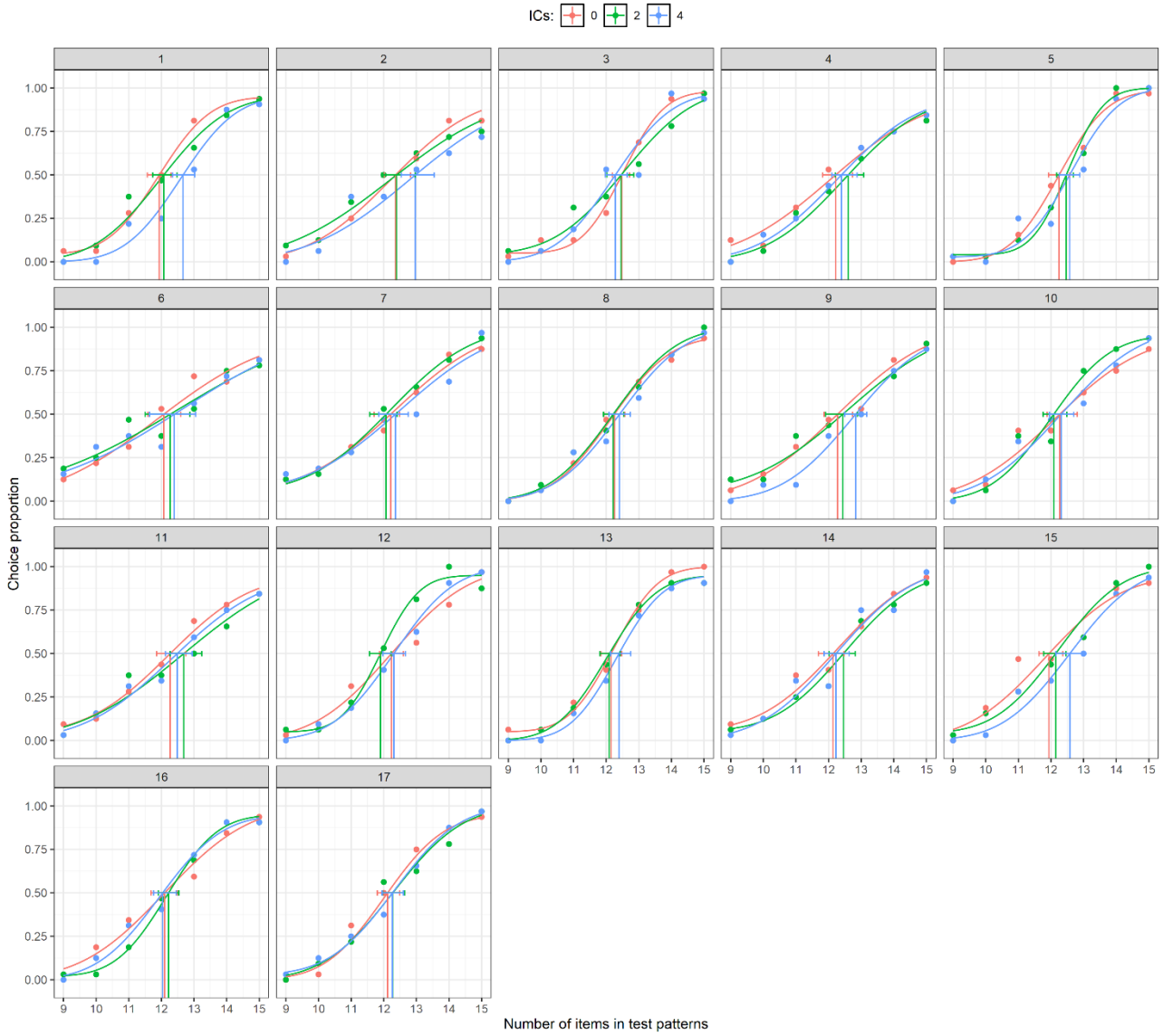


Figure S3.3: Individual psychometric functions for Experiment 1.

Participant	ICs	deviance	p
1	0	1.449	0.995
	2	4.553	0.79
	4	6.874	0.915
2	0	2.169	0.965
	2	2.584	0.995
	4	9.732	0.895
3	0	3.179	0.985
	2	4.071	1
	4	10.934	0.54
4	0	3.349	0.91
	2	4.144	0.87
	4	3.972	0.97
5	0	2.575	0.935
	2	5.297	0.795
	4	13.089	0.35
6	0	2.281	0.985
	2	3.654	0.835
	4	3.890	0.985
7	0	1.327	1
	2	0.829	1
	4	6.879	0.72
8	0	1.575	0.995
	2	5.125	0.99
	4	3.762	0.975
9	0	1.730	1
	2	3.805	0.83
	4	3.291	0.895
10	0	3.524	0.97
	2	6.985	0.96
	4	5.561	0.8
11	0	1.031	1
	2	4.109	0.87
	4	1.913	1
12	0	4.271	0.85
	2	7.095	0.71
	4	2.192	1
13	0	1.407	0.99
	2	1.573	0.985
	4	2.674	0.935
14	0	2.057	0.995
	2	2.054	0.97
	4	7.267	0.68
15	0	9.920	0.22
	2	5.622	0.605
	4	5.691	0.665
16	0	2.401	0.985
	2	1.497	0.985
	4	4.067	0.89
17	0	5.722	0.765
	2	5.345	0.94
	4	1.282	1

Table S3.4: Goodness of fit of psychometric functions was performed with Quickpsy. All *p* values were > .05, which suggests a good fit of the individual psychometric functions to the data.

Supplementary Results: Experiment 2A

Also in this case, we ran two separate Bayesian repeated-measures ANOVAs respectively on the PSE and CoV, with the number of aligned closed-inducers pairs as independent variable. For the PSE, the effect of the number of aligned closed-inducers pairs (this variable is called "ICs" in the main analyses and in the table below) showed a BF_{10} = .39, which suggests that data were 2.56 times more likely to occur under the null hypothesis than under the alternative hypothesis (Table S3.5).

For the CoV, the effect of the number of aligned closed-inducers pairs ("ICs") suggested that the evidence in favor of the null hypothesis was only 1.20 more likely (BF_{10} = 0.84) compared to the alternative hypothesis (Table S3.6). As for the supplementary analyses of Experiment 1, we correlated the individual CoV across the three different conditions. Pairwise correlations between IC conditions showed a positive and large relationship (all p < .001), suggesting that each individual CoV was stable across illusory conditions (Figure S3.5 and Table S3.7).

Table S3.5. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.719	2.564	1.000	
"ICs"	0.500	0.281	0.390	0.390	0.773

Note. All models include subject

Table S3.6. Bayesian ANOVA on the CoV.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.543	1.190	1.000	
"ICs"	0.500	0.457	0.840	0.840	0.486

Note. All models include subject

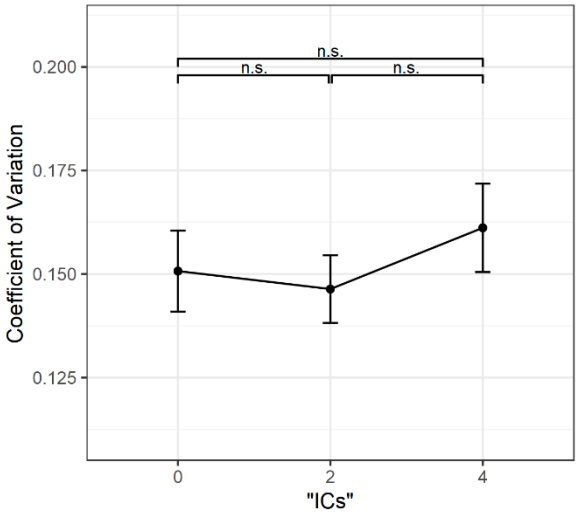


Figure S3.4: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

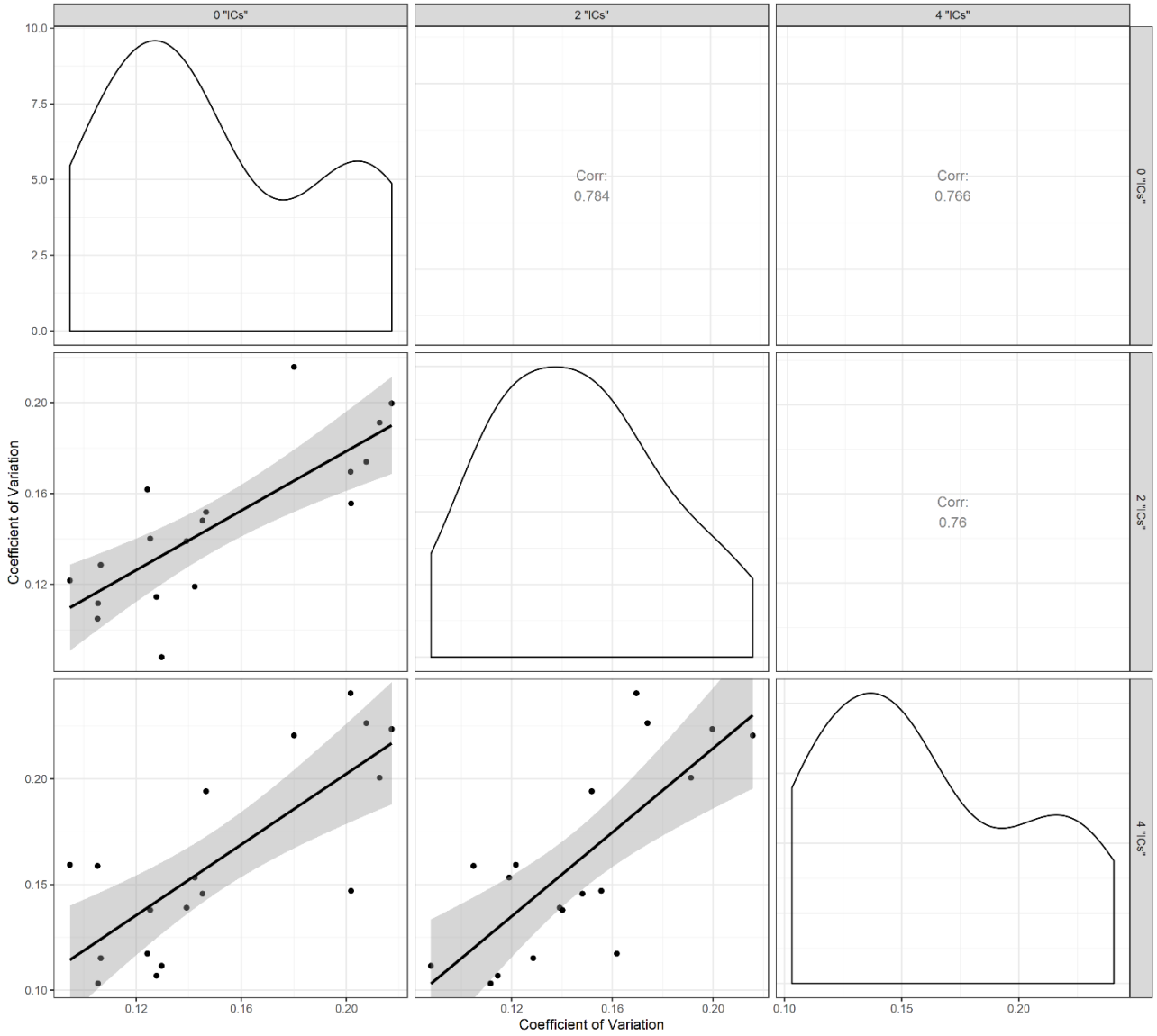


Figure S3.5: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Table S3.7. CoV correlation values among conditions.

Pearson Correlations				
		0 "ICs"	**2 "ICs"**	**4 "ICs"**
0 "ICs"	Pearson's r	—		
	p-value	—		
2 "ICs"	Pearson's r	0.784 ***	—	
	p-value	< .001	—	
4 "ICs"	Pearson's r	0.766 ***	0.760 ***	—
	p-value	< .001	< .001	—

* p < .05, ** p < .01, *** p < .001

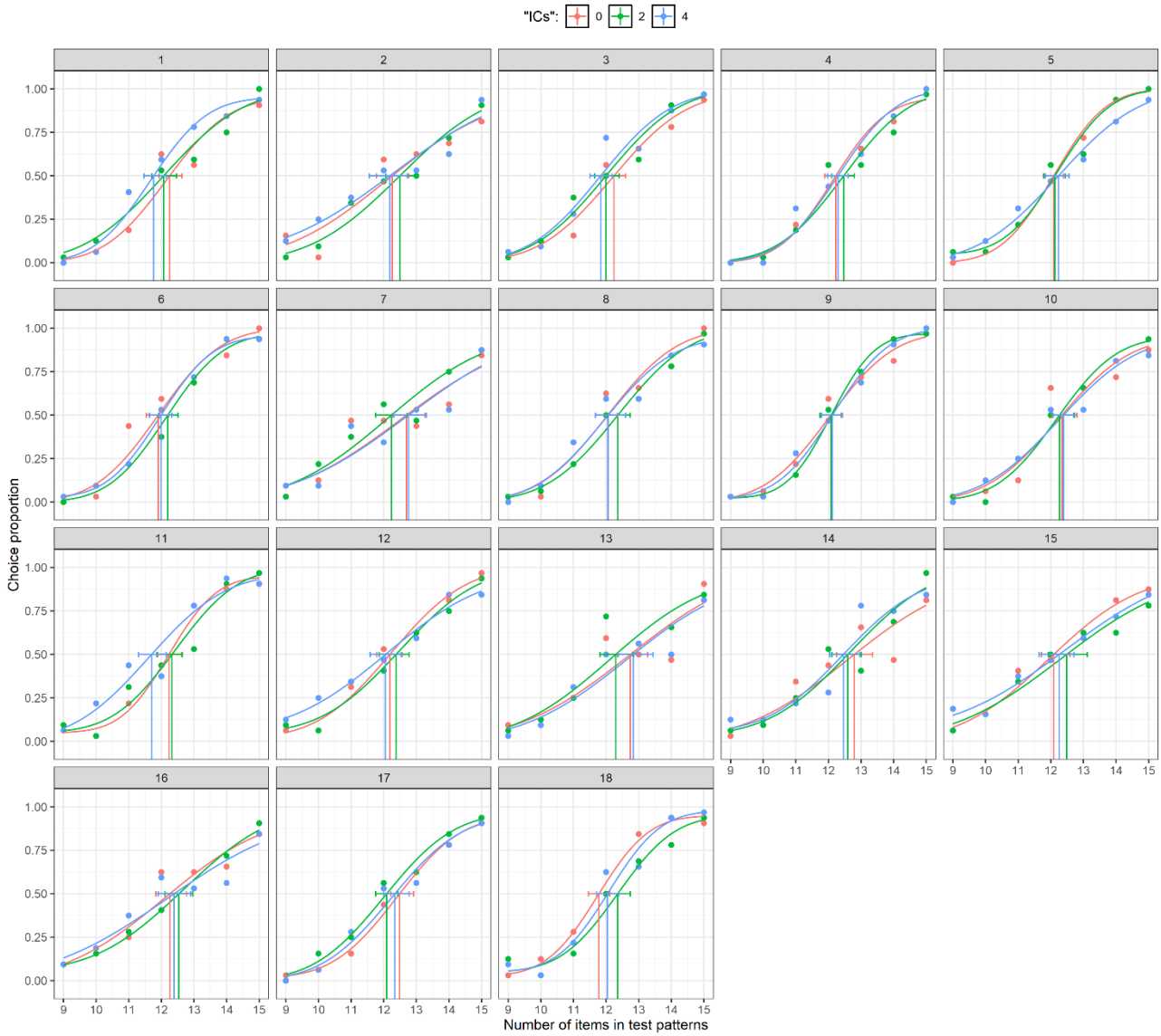


Figure S3.6: Individual psychometric functions for the Experiment 2A.

Participant	"ICs"	deviance	p
1	0	7.799	0.51
	2	9.958	0.68
	4	5.510	0.715
2	0	11.378	0.525
	2	4.313	0.9
	4	6.439	0.665
3	0	3.917	0.87
	2	4.001	0.855
	4	6.875	0.935
4	0	8.187	0.72
	2	8.130	0.435
	4	10.366	0.37
5	0	1.738	0.99
	2	4.842	0.985
	4	1.254	1
6	0	12.211	0.475
	2	3.332	0.965
	4	1.058	1
7	0	12.334	0.54
	2	7.237	0.69
	4	10.319	0.165
8	0	11.546	0.315
	2	2.664	0.96
	4	6.493	0.645
9	0	3.602	0.895
	2	1.050	1
	4	3.473	0.985
10	0	11.093	0.215
	2	6.295	0.645
	4	5.830	0.92
11	0	4.039	0.845
	2	7.780	0.73
	4	7.815	0.675
12	0	4.066	0.95
	2	4.000	0.945
	4	1.718	0.98
13	0	15.381	0.185
	2	12.043	0.255
	4	8.456	0.395
14	0	11.336	0.305
	2	10.735	0.65
	4	8.217	0.725
15	0	2.089	0.995
	2	3.013	0.98
	4	1.627	1
16	0	5.483	0.915
	2	1.014	1
	4	6.967	0.855
17	0	0.732	1
	2	4.584	0.885
	4	5.170	0.9
18	0	2.077	0.975
	2	6.362	0.88
	4	7.066	0.695

Table S3.8: Goodness of fit of psychometric functions was performed with Quickpsy. All *p* values were > .05, which suggests a good fit of the individual psychometric functions to the data.

Supplementary Results: Experiment 2B

As in the Experiment 1 and 2A, we also ran two separated one-way repeated measures Bayesian ANOVAs with the number of "ICs" as independent factor and the PSEs or the CoV as dependent variable. Results of the analysis of PESs indicate a moderate evidence in favor of the null hypothesis (e.g., no difference across "ICs" conditions), since this hypothesis was 3.92 times more likely (BF_{10} = 0.255) than the alternative hypothesis (Table S3.9). A similar Bayesian ANOVA was run over the data with the CoV as dependent variable, suggesting that null hypothesis (e.g., no difference across "ICs" conditions) was 2.63 times more likely (BF_{10} = 0.38) than the alternative hypothesis (Table S10). Pairwise correlations between IC conditions showed a positive and large relationship (all $p < .001$), suggesting that each individual CoV was stable across illusory conditions (Figure S3.8 and Table S3.11).

Table S3.9. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.797	3.915	1.000	
"ICs"	0.500	0.203	0.255	0.255	0.665

Note. All models include subject

Table S3.10. Bayesian ANOVA on the CoV.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.725	2.637	1.000	
"ICs"	0.500	0.275	0.379	0.379	0.497

Note. All models include subject

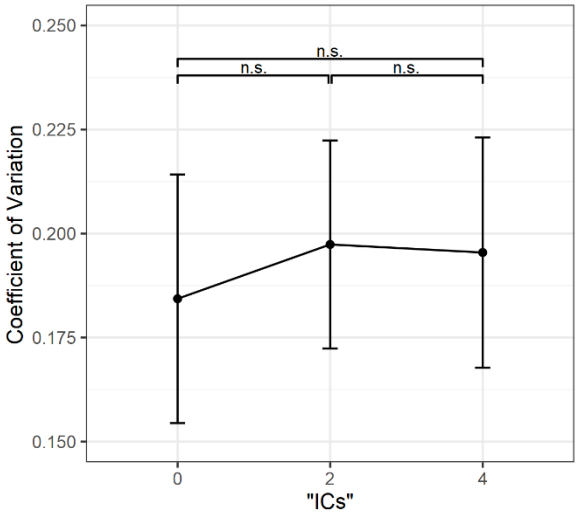


Figure S3.7: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

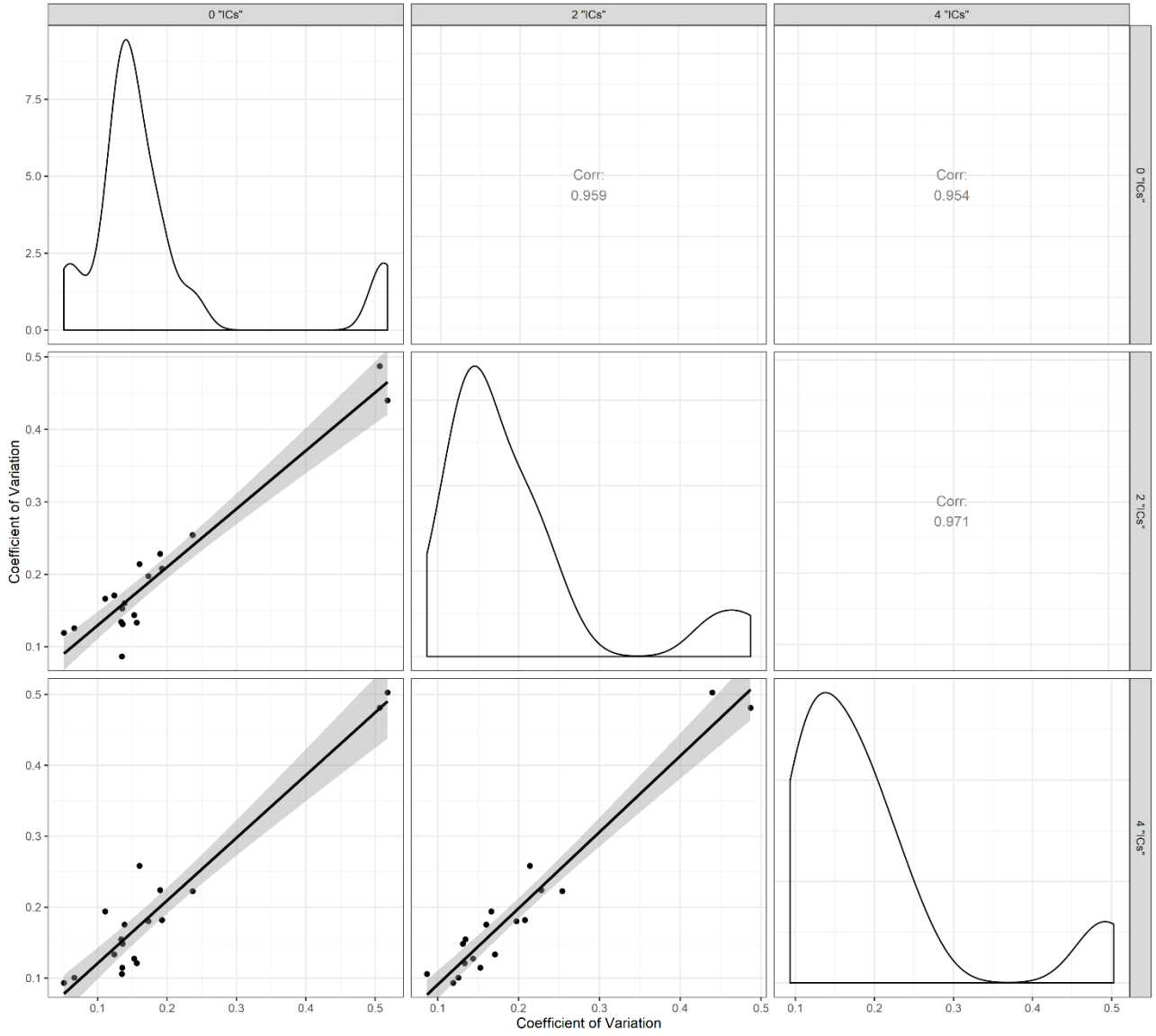


Figure S3.8: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Table S3.11. CoV correlation values among conditions.

Pearson Correlations				
		0 "ICs"	**2 "ICs"**	**4 "ICs"**
0 "ICs"	Pearson's r	—		
	p-value	—		
2 "ICs"	Pearson's r	0.959 ***	—	
	p-value	< .001	—	
4 "ICs"	Pearson's r	0.954 ***	0.971 ***	—
	p-value	< .001	< .001	—

* p < .05, ** p < .01, *** p < .001

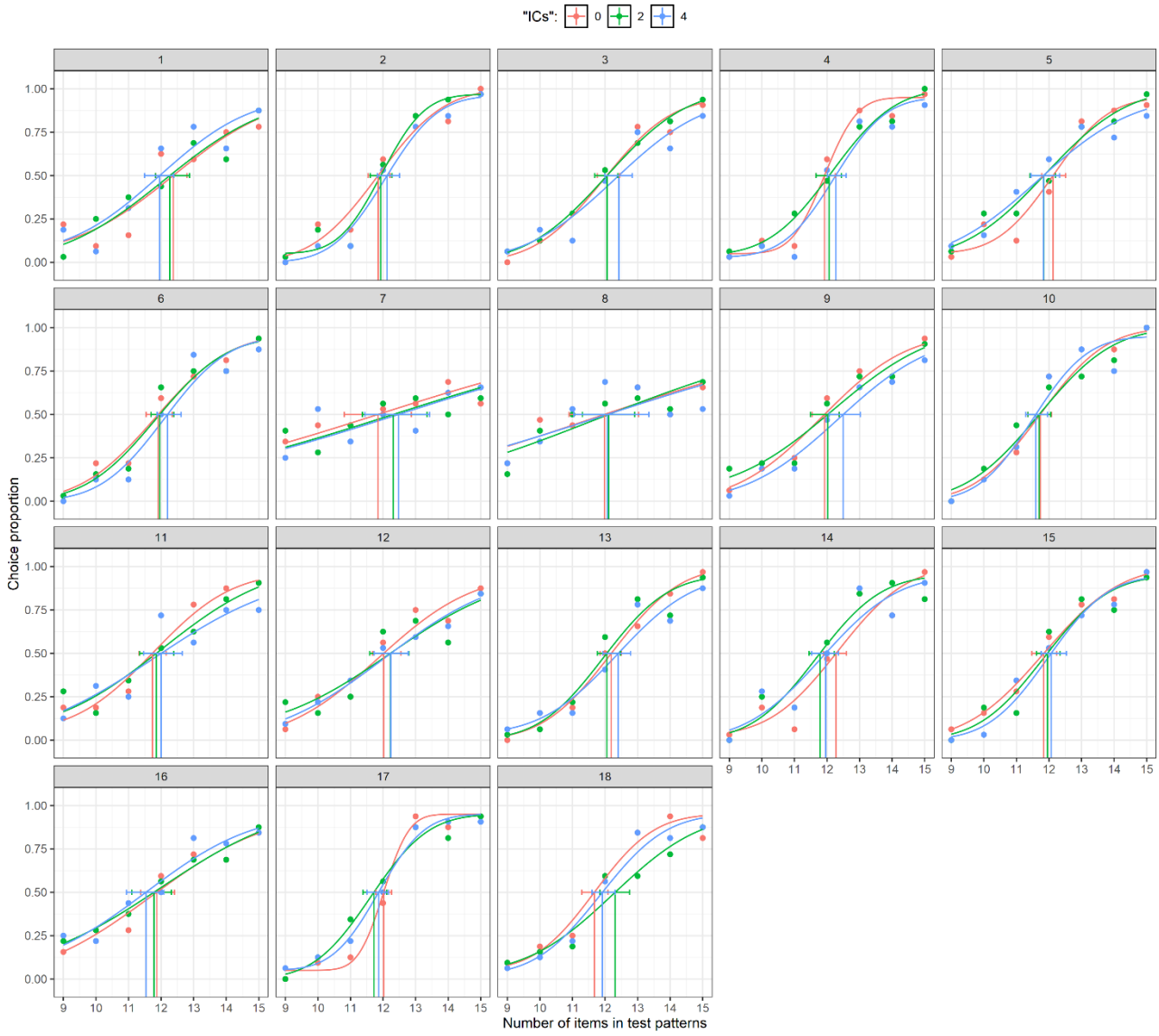


Figure S3.9: Individual psychometric functions for the Experiment 2B.

Participant	"ICs"	deviance	p
1	0	13.358	0.08
	2	8.050	0.55
	4	15.016	0.15
2	0	12.899	0.135
	2	7.962	0.95
	4	6.074	0.99
3	0	5.813	0.845
	2	0.557	1
	4	9.050	0.4
4	0	8.749	0.94
	2	4.553	0.98
	4	13.552	0.575
5	0	9.033	0.3
	2	4.725	0.765
	4	4.982	0.875
6	0	4.089	0.85
	2	7.685	0.585
	4	11.579	0.33
7	0	2.847	0.915
	2	5.351	0.84
	4	7.453	0.68
8	0	6.378	0.725
	2	6.274	0.57
	4	13.500	0.075
9	0	5.442	0.64
	2	5.250	0.715
	4	3.697	0.985
10	0	9.829	0.42
	2	11.346	0.89
	4	16.226	0.59
11	0	3.214	0.935
	2	5.275	0.725
	4	10.794	0.19
12	0	5.654	0.66
	2	11.278	0.915
	4	1.723	0.97
13	0	4.142	0.865
	2	8.421	0.37
	4	7.208	0.49
14	0	17.978	0.385
	2	15.188	0.135
	4	18.082	0.29
15	0	2.393	0.95
	2	12.531	0.18
	4	7.653	0.475
16	0	3.824	0.9
	2	1.591	0.985
	4	4.686	0.675
17	0	8.294	0.9
	2	6.205	0.965
	4	2.793	0.945
18	0	8.756	0.345
	2	4.876	0.79
	4	5.439	0.94

Table S3.12: Goodness of fit of psychometric functions was performed with Quickpsy. All *p* values were > .05, which suggests a good fit of the individual psychometric functions to the data.

Supplementary Results: Experiment 3

We ran two separate Bayesian repeated-measures ANOVAs (2 x 3) respectively on the PSE and CoV, with the number of ICs and the convex hull size as independent variables. For the PSE, the model with both main effects, but not the interaction, received strong evidence in favor of the alternative hypothesis (BF_{10} = 243.9), compared to the other models (Table S3.13).

For the CoV, the model containing only the main effect of the convex hull size received strong evidence in favor of the alternative hypothesis (BF_{10} = 914.7) compared to the other models (Table S3.14). Finally, we correlated the individual CoV across the different conditions. Pairwise correlations among conditions showed a positive and large relationship (all p < .001), suggesting that each individual CoV was stable across experimental conditions (Figure S3.11 and Table S3.15).

Table S3.13. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.003	0.014	1.000	
Convex Hull + ICs	0.200	0.822	18.448	243.943	6.710
Convex Hull + ICs + Convex Hull ✻ ICs	0.200	0.083	0.362	24.639	8.884
Convex Hull	0.200	0.055	0.233	16.342	0.920
ICs	0.200	0.037	0.153	10.911	1.258

Note. All models include subject

Table S3.14. Bayesian ANOVA on the CoV.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	8.403e -4	0.003	1.000	
Convex Hull	0.200	0.769	13.285	914.704	7.256
Convex Hull + ICs	0.200	0.210	1.066	250.509	1.920
Convex Hull + ICs + Convex Hull ✻ ICs	0.200	0.020	0.081	23.637	1.852
ICs	0.200	2.139e -4	8.559e -4	0.255	0.745

Note. All models include subject

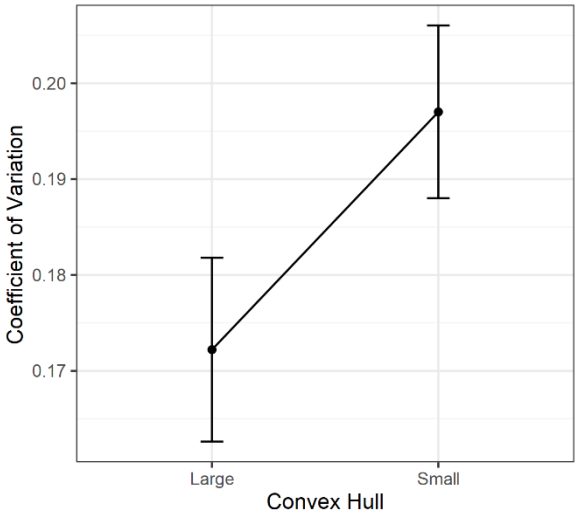


Figure S3.10: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

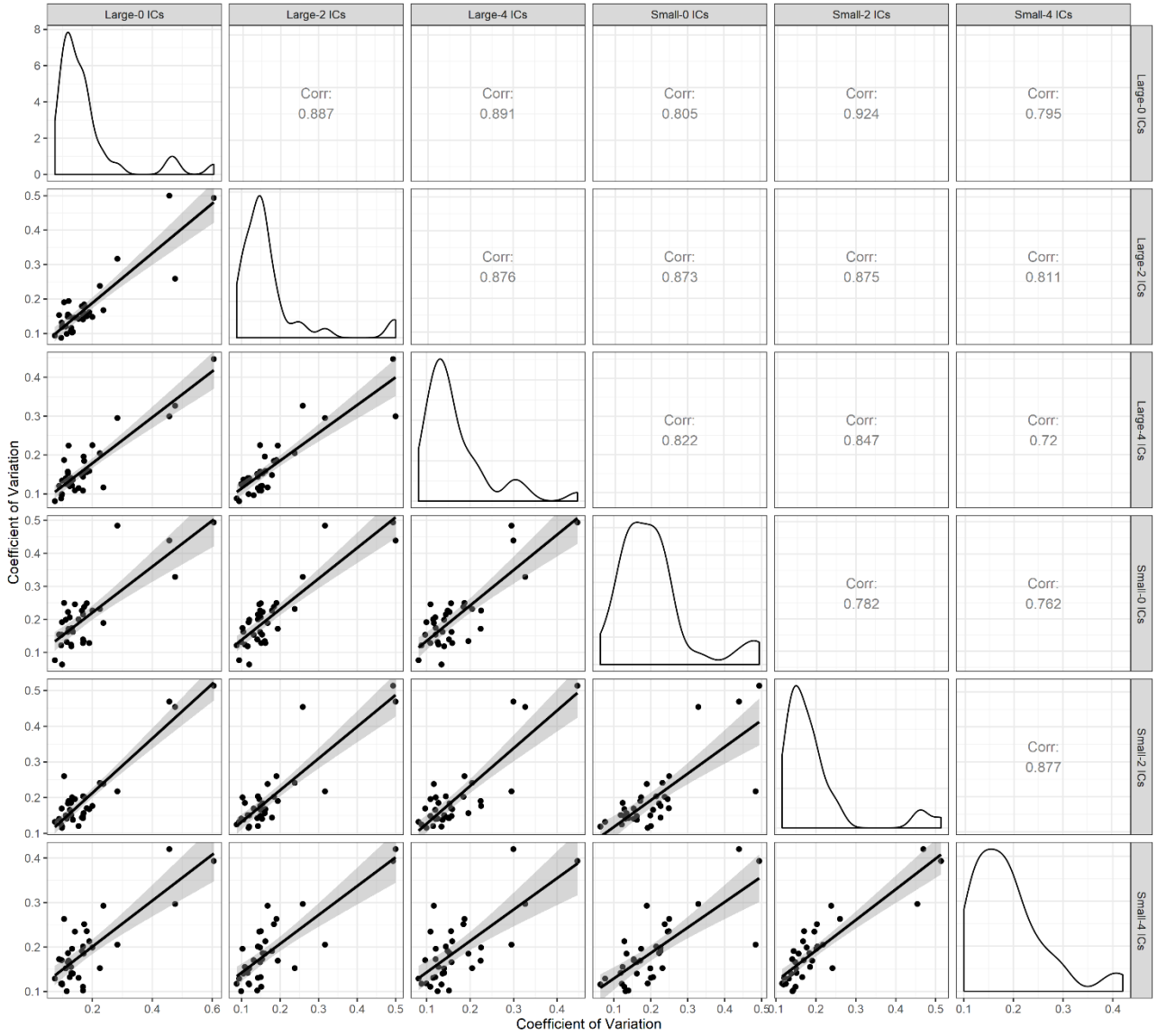


Figure S3.11: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Table S3.15. CoV correlation values among conditions.

Pearson Correlations

		Large-0 ICs	Large-2 ICs	Large-4 ICs	Small-0 ICs	Small-2 ICs	Small-4 ICs
Large-0 ICs	Pearson's r	—					
	p-value	—					
Large-2 ICs	Pearson's r	0.887 ***	—				
	p-value	< .001	—				
Large-4 ICs	Pearson's r	0.891 ***	0.876 ***	—			
	p-value	< .001	< .001	—			
Small-0 ICs	Pearson's r	0.805 ***	0.873 ***	0.822 ***	—		
	p-value	< .001	< .001	< .001	—		
Small-2 ICs	Pearson's r	0.924 ***	0.875 ***	0.847 ***	0.782 ***	—	
	p-value	< .001	< .001	< .001	< .001	—	
Small-4 ICs	Pearson's r	0.795 ***	0.811 ***	0.720 ***	0.762 ***	0.877 ***	—
	p-value	< .001	< .001	< .001	< .001	< .001	—

* p < .05, ** p < .01, *** p < .001

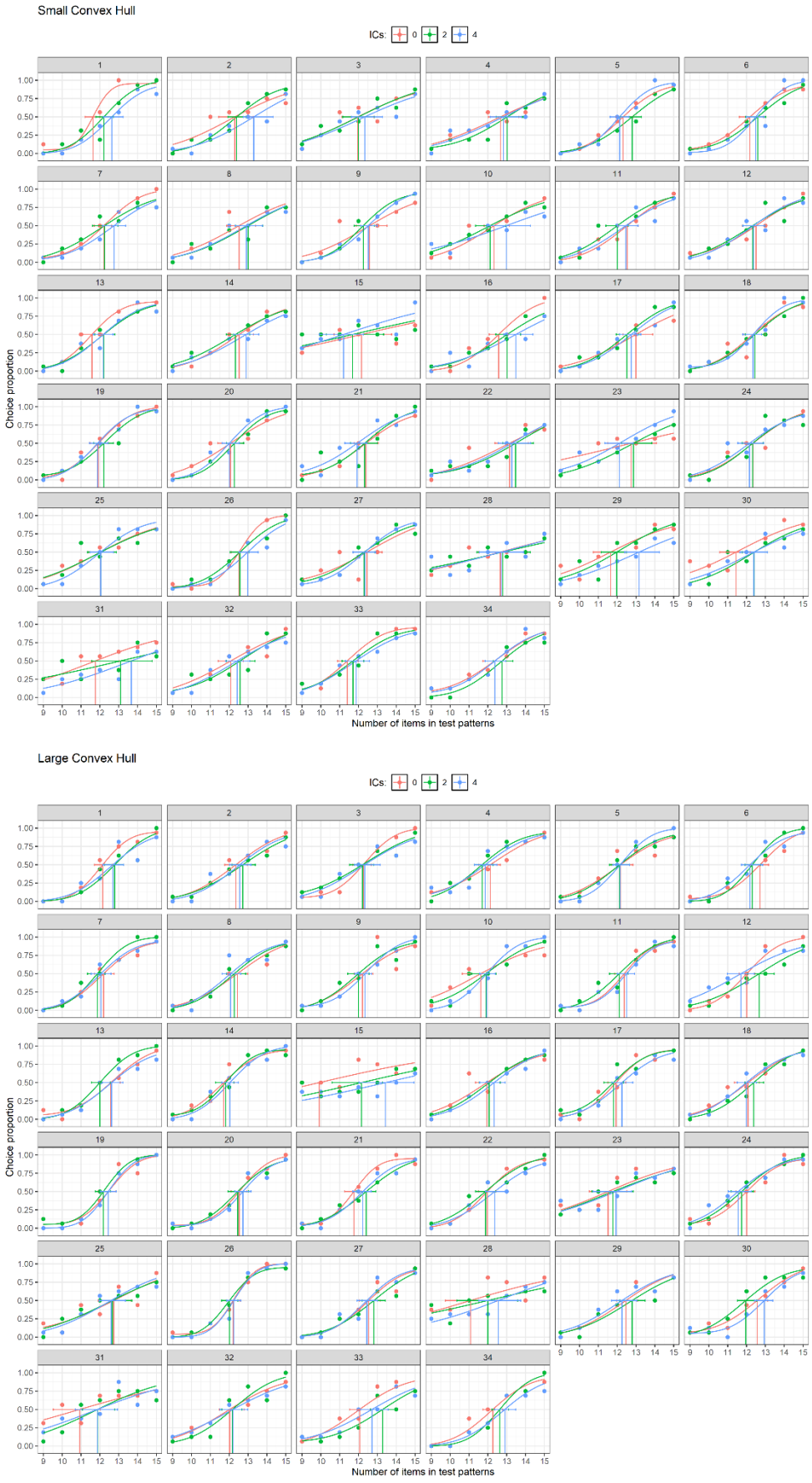


Figure S3.12: Individual psychometric functions for the Experiment 3.

Participant	ICs	Convex Hull	deviance	p
1	0	Large	3.343	1
		Small	10.437	0.935
	2	Large	9.376	0.465
		Small	8.271	0.56
	4	Large	10.594	0.28
		Small	3.915	0.9
2	0	Large	3.000	0.925
		Small	8.525	0.36
	2	Large	1.606	1
		Small	3.493	0.925
	4	Large	8.364	0.42
		Small	6.593	0.595
3	0	Large	2.012	0.995
		Small	7.152	0.465
	2	Large	2.691	0.98
		Small	4.102	0.84
	4	Large	5.011	0.95
		Small	5.239	0.79
4	0	Large	1.273	1
		Small	3.830	0.995
	2	Large	4.411	0.905
		Small	4.982	0.81
	4	Large	5.415	0.84
		Small	7.228	0.915
5	0	Large	7.813	0.595
		Small	1.203	1
	2	Large	2.464	0.99
		Small	2.627	0.98
	4	Large	5.246	0.88
		Small	6.552	0.815
6	0	Large	6.013	0.855
		Small	0.528	1
	2	Large	3.534	0.915
		Small	1.581	1
	4	Large	4.163	0.865
		Small	11.260	0.165
7	0	Large	2.686	0.965
		Small	2.583	0.98
	2	Large	4.992	0.93
		Small	6.264	0.555
	4	Large	3.008	0.94
		Small	2.803	0.965
8	0	Large	3.431	0.935
		Small	6.161	0.805
	2	Large	2.960	1
		Small	7.640	0.735
	4	Large	11.410	0.83
		Small	2.752	0.99
9	0	Large	19.969	0.01
		Small	9.797	0.305
	2	Large	7.501	0.725
		Small	2.007	1
	4	Large	2.396	0.955
		Small	1.335	1
10	0	Large	8.855	0.395
		Small	5.226	0.99
	2	Large	4.678	0.785
		Small	1.304	0.99
	4	Large	3.406	0.925
		Small	3.516	0.965

11	0	Large	6.344	0.655
		Small	2.955	0.995
	2	Large	5.262	0.735
		Small	4.191	0.895
	4	Large	6.030	0.9
		Small	7.407	0.515
12	0	Large	5.005	0.89
		Small	2.479	0.94
	2	Large	3.270	0.94
		Small	6.041	0.885
	4	Large	4.076	0.875
		Small	5.738	0.68
13	0	Large	5.628	0.66
		Small	3.698	0.96
	2	Large	2.405	0.975
		Small	6.245	0.85
	4	Large	4.104	1
		Small	6.936	0.715
14	0	Large	5.348	0.905
		Small	2.781	0.96
	2	Large	2.885	0.96
		Small	1.020	0.995
	4	Large	7.305	0.9
		Small	3.432	0.905
15	0	Large	5.752	0.975
		Small	6.565	0.55
	2	Large	4.205	0.79
		Small	7.203	0.605
	4	Large	3.481	0.955
		Small	7.516	0.75
16	0	Large	10.454	0.335
		Small	9.140	0.415
	2	Large	0.198	1
		Small	1.659	0.99
	4	Large	3.385	0.93
		Small	8.568	0.765
17	0	Large	7.262	0.59
		Small	2.301	0.955
	2	Large	5.230	0.76
		Small	4.057	1
	4	Large	4.016	0.91
		Small	3.724	0.885
18	0	Large	4.061	0.86
		Small	9.333	0.435
	2	Large	3.369	0.98
		Small	5.237	0.765
	4	Large	2.343	0.96
		Small	5.720	0.915
19	0	Large	9.319	0.485
		Small	5.763	0.89
	2	Large	2.021	0.985
		Small	7.319	0.915
	4	Large	1.660	0.98
		Small	4.841	0.935
20	0	Large	3.929	0.87
		Small	4.130	0.955
	2	Large	2.391	0.99
		Small	1.993	0.995
	4	Large	5.623	0.795
		Small	4.277	0.8
21	0	Large	4.010	0.94
		Small	0.774	1

	2	Large	2.310	0.965
		Small	16.009	0.05
	4	Large	4.938	0.75
		Small	6.176	0.695
22	0	Large	4.369	0.94
		Small	10.175	0.29
	2	Large	5.735	0.715
		Small	3.715	0.925
	4	Large	5.837	0.7
		Small	2.997	1
23	0	Large	6.378	0.64
		Small	4.192	0.845
	2	Large	2.643	1
		Small	0.414	1
	4	Large	5.401	0.65
		Small	3.831	0.965
24	0	Large	7.310	0.745
		Small	5.666	0.865
	2	Large	2.176	0.985
		Small	12.455	0.385
	4	Large	5.235	0.785
		Small	6.150	0.735
25	0	Large	10.889	0.35
		Small	2.124	1
	2	Large	5.894	0.86
		Small	7.441	0.91
	4	Large	4.610	0.86
		Small	3.677	1
26	0	Large	3.540	0.975
		Small	2.453	1
	2	Large	0.708	1
		Small	5.355	0.705
	4	Large	4.329	0.625
		Small	4.624	0.775
27	0	Large	7.205	0.895
		Small	11.67	0.455
	2	Large	4.756	0.965
		Small	4.749	0.735
	4	Large	8.558	0.385
		Small	0.511	1
28	0	Large	10.69	0.22
		Small	4.758	0.76
	2	Large	4.336	0.915
		Small	2.949	0.915
	4	Large	1.864	0.985
		Small	7.949	0.41
29	0	Large	5.187	0.715
		Small	6.246	0.66
	2	Large	8.981	0.61
		Small	7.160	0.82
	4	Large	7.106	0.58
		Small	1.724	0.985
30	0	Large	2.149	1
		Small	7.157	0.75
	2	Large	4.822	0.88
		Small	5.361	0.645
	4	Large	7.439	0.535
		Small	4.644	0.765
31	0	Large	4.682	0.805
		Small	3.030	0.955
	2	Large	9.610	0.575
		Small	7.322	0.42

	4	Large	7.793	0.565
		Small	4.386	0.845
32	0	Large	2.304	0.99
		Small	6.084	0.56
	2	Large	11.487	0.875
		Small	7.798	0.43
	4	Large	0.805	1
		Small	4.127	0.93
33	0	Large	6.006	0.98
		Small	2.792	0.925
	2	Large	6.009	0.895
		Small	4.208	0.875
	4	Large	4.026	0.86
		Small	1.202	1
34	0	Large	6.671	0.945
		Small	1.896	1
	2	Large	4.744	0.865
		Small	6.177	0.735
	4	Large	5.450	0.965
		Small	4.854	0.8

Table S3.16: Goodness of fit of psychometric functions was performed with Quickpsy. Except for the subject 9 ($p < .01$), all p values were > .05, suggesting a good fit of the individual psychometric functions to the data.

Supplementary Results: Experiment 4

As for the previous experiments, we ran two separate Bayesian repeated-measures ANOVAs for the subjective estimation index and for the CoV, with the target Numerosity and the number of ICs as independent factors. For this analysis, we compared the complex model, which include all the possible combinations of main effects and interactions, against the null model. For the subjective estimation, we found that the model including only the two main effects (but not the interaction) showed very strong evidence (BF10 > 100) compared to the null model (Table S3.17).

For the CoV, the data provided strong support for the full model including all the main effects and the interaction. Indeed, for this model we found a BF_{10} = .014, which suggests data were 70 times more likely to occur with this model under the null hypothesis (e.g., constant CoV across conditions) than under the alternative hypothesis (Table S3.18). Finally, individual CoV were significantly and positively correlated across IC conditions (all p < .001), hence suggesting a stable individual CoV across illusory conditions (Figure S3.13 and Table S3.19).

Table S3.17. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	3.323e -172	1.329e -171	1.000	
ICs + Numerosity	0.200	0.970	128.091	2.919e +171	1.145
ICs + Numerosity + ICs ✻ Numerosity	0.200	0.018	0.071	5.281e +169	32.855
Numerosity	0.200	0.013	0.052	3.833e +169	0.482
ICs	0.200	2.333e -173	9.333e -173	0.070	0.687

Note. All models include subject

Table S3.18. Bayesian ANOVA on the CoV.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.252	1.346	1.000	
Numerosity	0.200	0.638	7.043	2.533	0.364
Numerosity + ICs	0.200	0.078	0.339	0.311	1.020
ICs	0.200	0.029	0.118	0.114	1.083
Numerosity + ICs + Numerosity ✻ ICs	0.200	0.004	0.014	0.014	2.856

Note. All models include subject

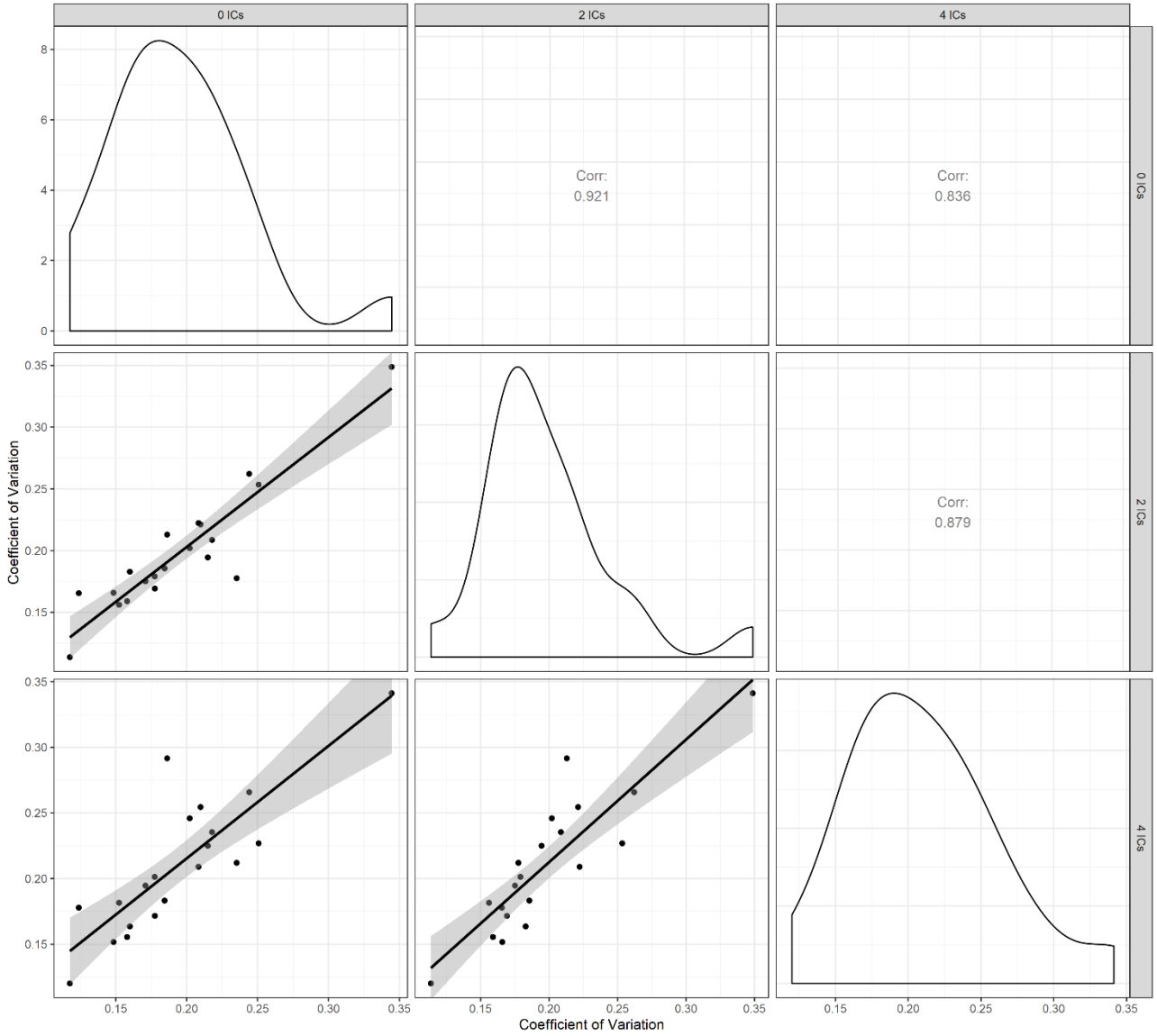


Figure S3.13: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Table S3.19. CoV correlation values among conditions.

Pearson Correlations

		0 ICs	2 ICs	4 ICs
0 ICs	Pearson's r	—		
	p-value	—		
2 ICs	Pearson's r	0.921 ***	—	
	p-value	< .001	—	
4 ICs	Pearson's r	0.836 ***	0.879 ***	—
	p-value	< .001	< .001	—

* p < .05, ** p < .01, *** p < .001

Simulation

We ran a Python simulation to demonstrate the possible effects of the different constrains applied over the distance between items inducers, affecting the overall spatial distribution of the items. In particular, we generated the stimuli as in the original paper of K&M2016 and we calculated the area of convex hull (in px^2) for each stimulus. We resampled a series of fictitious stimuli for the condition in which the test stimuli contained 12 items, respectively for the 0 ICs, 2 ICs, and 4 ICs condition. We generated 10 random samples (n = 10), each sample formed by a subset of 8 random patterns for each ICs level (that were repeated across trials in the experiment), with the same constrains as used in the K&M2016. Thus, the patterns for the 0 ICs condition were generated with the only constrain imposed to the spatial disposition of the inducers to not overlap (minimum interdot distance: 22 px, center-to-center). On the other hand, when the 4 ICs stimuli were generated, 4 pairs of inducers (8 dots total in each pattern) were constrained in order to be neighbor inducers (e.g., within a random distance of 22, 25, 28 or 31 px). Similarly, 2 pairs of inducers (4 dots in each pattern) were constrained for the 2 ICs condition.

For each resample we calculated the mean area of the convex hull of the set of 8 patterns, for each ICs level. The simulated data relative to the mean area of the generated stimuli for each ICs level were analyzed with an independent-samples Bayesian ANOVA.

Results showed that the alternative hypothesis (e.g., the mean convex hull area of the stimuli set varied with the number of ICs) received a strong support (BF_{10} = 488.95) compared to the null hypothesis (e.g., no difference in mean convex hull area across ICs conditions). Indeed, the mean area of the convex hull clearly decreased with the number of ICs in the stimuli set. This strongly suggests that the method of K&M might produce biased stimuli in which the mean convex hull area of the set actually decreases with the number of ICs. Thus, a simple decrease in convex hull area rather than the segmentation process itself may explain the results by K&M.

Results

	Condition		
n	**0 ICs**	**2 ICs**	**4 ICs**
1	18320.625	16749.3125	16351.625
2	17776.5	18216.0625	17483.0625
3	16570.5	18142.8125	16294.25
4	18331.875	16584.375	14056.375
5	19305.06	18885.0	16455.187
6	18292.68	16240.0625	14131.937
7	17404.75	15684.625	13677.75
8	19923.375	17559.0625	14836.812
9	18008.125	15649.0625	15857.875
10	18513.062	16186.9375	16005.9375
Mean Area (px^2)	18245	16990	15515
Sd	932	1139	1260

Table 1. Random data obtained with a Python script generating the stimuli as in K&M2016.

Model Comparison - Area

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.00204	0.00205	1.00	
ICs	0.500	0.99796	488.95344	488.95	1.62e-4

Table 2. Bayesian ANOVA.

Post Hoc Comparisons - ICs

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
2 ICs	4 ICs	0.587	2.48	4.22	6.89e-4
	0 ICs	0.587	2.30	3.92	6.12e-4
4 ICs	0 ICs	0.587	309.36	526.66	4.07e-6

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, r = 1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Table 3. Bayesian Post-Hoc Comparisons.

Chapter 4

Supplementary Results: Experiment 1

We ran two separate Bayesian repeated-measures ANOVAs (2 x 3) respectively on the PSE and CoV, with the number of ICs and the inducers color as independent variables. From an inspection of the table below (Table S4.1) for the PSE, the model that most outperforms the Null model is the one with the main effect of ICs only, which received strong evidence in favor of the alternative hypothesis (BF_{10} = 66.7). Furthermore, adding the main factor of color of inducers (BF_{10} = 21.2) or the interaction (BF_{10} = 3.95) makes the model less competitive. We also compared the strength of the Bayes factor against the Null model for the models that exclude or include the critical interaction term. The evidence *against* including the interaction is roughly a factor of six, compared to the model with the main factors. This can be obtained as 21.2 / 3.95 ≈ 5.36. Thus, the data are almost 6 times more likely under the two main effects model than under the full model (i.e., the one including also the interaction). Similarly, the data are 17 times more likely under the model with the main effect of ICs only, compared to the full model (66.7/3.95 ≈ 16.88). Finally, the main effect of color of inducers received substantial support in favor of the null hypothesis (BF_{10} = 0.311). In sum, the Bayesian ANOVA for the PSE reveals that the data provide very strong support for the main effect of ICs, but good evidence against the color of inducers. The data also provide good evidence against including the interaction term.

The analysis of CoV revealed that the full model, containing the main effects and the interaction, received strong evidence in favor of the alternative hypothesis (BF_{10} = 0.031) compared to all the other simple models (Table S4.2).

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.011	0.043	1.000	
ICs	0.200	0.716	10.074	66.721	0.760
Inducers color + ICs	0.200	0.228	1.180	21.232	1.699
Inducers color + ICs + Inducers color ✻ ICs	0.200	0.042	0.177	3.950	4.411
Inducers color	0.200	0.003	0.013	0.311	3.751

Note. All models include subject

Table S4.1. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	0.543	4.756	1.000	
ICs	0.200	0.277	1.529	0.509	1.852
Inducers Color	0.200	0.109	0.487	0.200	2.033
Inducers Color + ICs	0.200	0.055	0.233	0.101	1.966
Inducers Color + ICs + Inducers Color ✻ ICs	0.200	0.017	0.068	0.031	2.132

Note. All models include subject

Table S4.2. Bayesian ANOVA on the CoV.

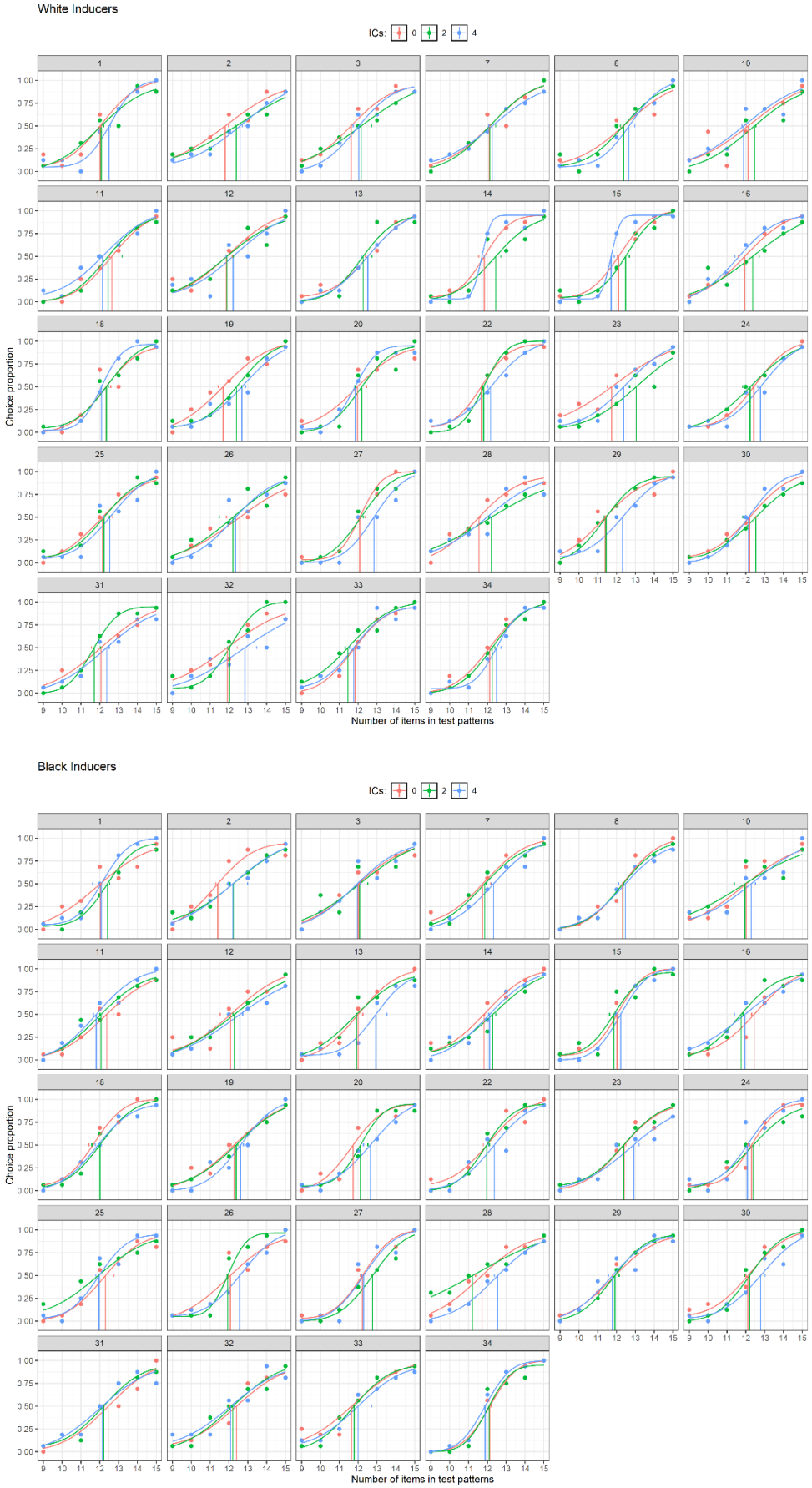


Figure S4.1: Individual psychometric functions for Experiment 1.

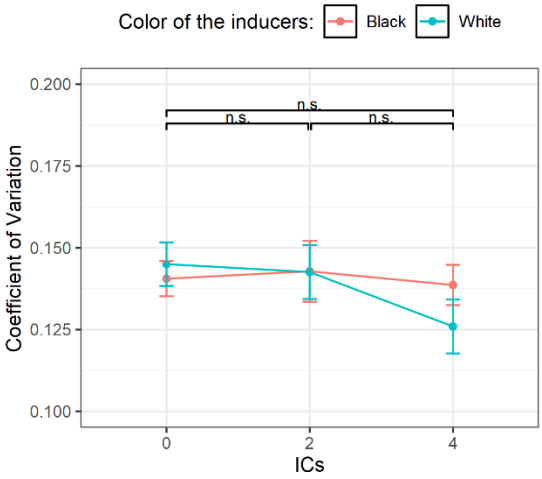


Figure S4.2: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

Supplementary Results: Experiment 2

As in the Experiment 1 we ran two separate one-way Bayesian repeated-measures ANOVAs respectively on the PSE and CoV, with the number of ICs as independent variable. The analysis of PSE (Table S4.3) revealed that the factor number of ICs received overwhelming evidence in favor of the alternative hypothesis (BF_{10} = 707.1). Furthermore, the analysis of the CoV (Table S4) suggested that the factor number of ICs received anecdotal evidence in favor of the null hypothesis (BF_{10} = 0.399).

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.001	0.001	1.000	
ICs	0.500	0.999	707.102	707.102	1.674

Note. All models include subject

Table S4.3. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.715	2.504	1.000	
ICs	0.500	0.285	0.399	0.399	1.271

Note. All models include subject

Table S4.4. Bayesian ANOVA on the CoV.

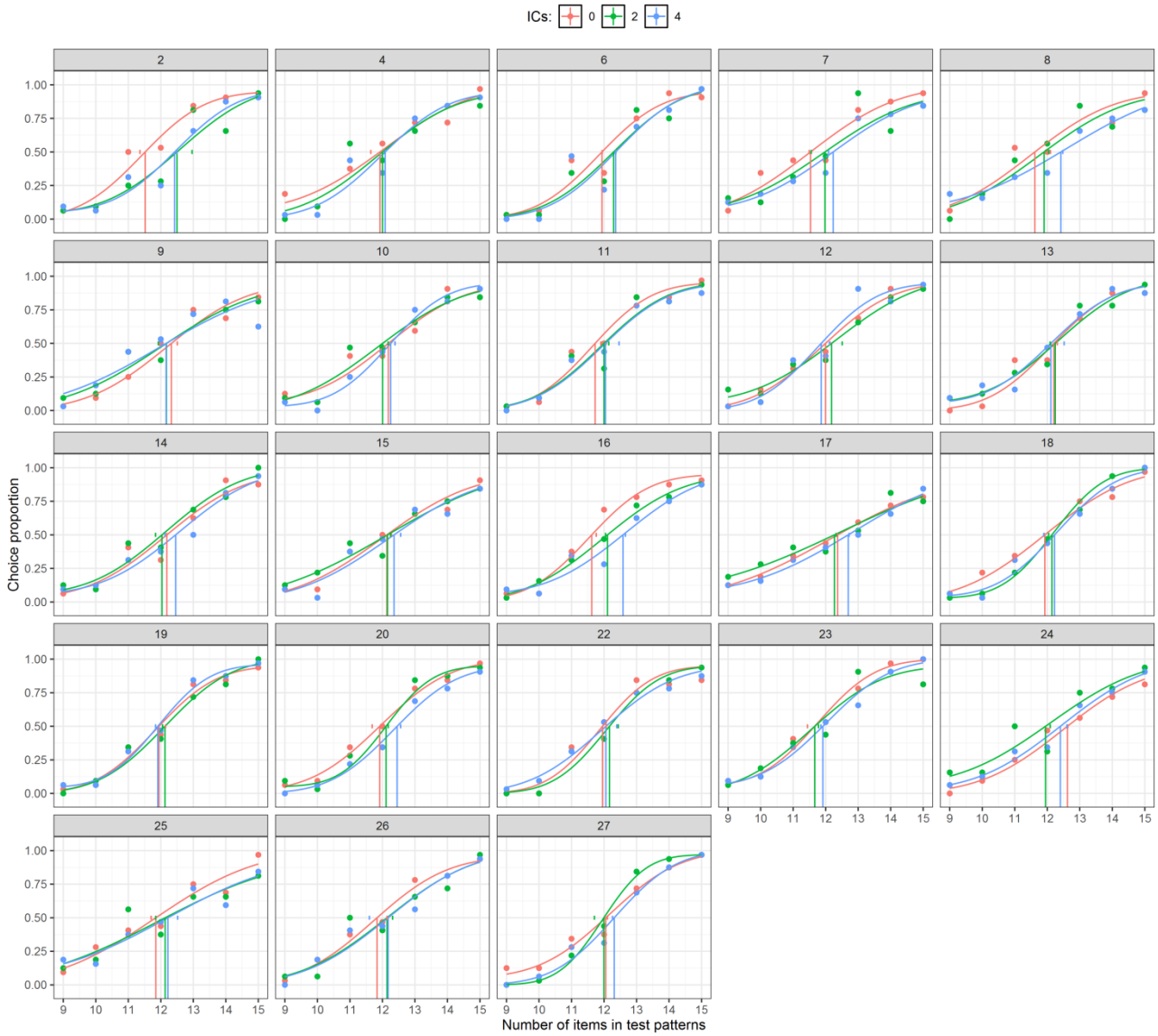


Figure S4.3: Individual psychometric functions for Experiment 2.

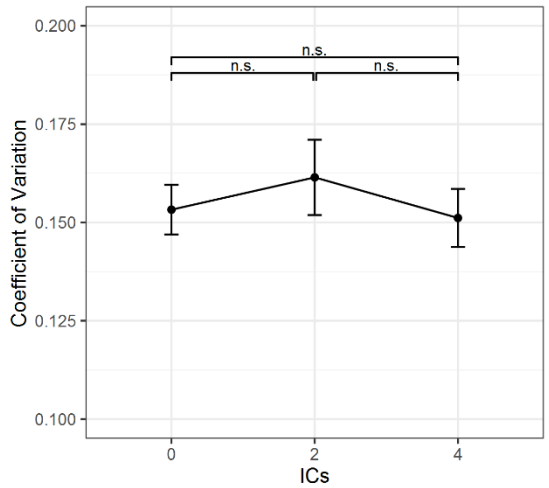


Figure S4.4: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

Supplementary Results: Experiment 3

As in the previous experiments, we ran two separate one-way Bayesian repeated-measures ANOVAs respectively on the PSE and CoV, with the number of ICs as independent variable, as in the second experiment. The analysis of PSE (Table S4.5) revealed that the factor number of ICs received only anecdotal evidence in favor of the null hypothesis (BF_{10} = 0.41). Finally, the analysis of the CoV (Table S4.6) suggested that the factor number of ICs received substantial evidence in favor of the null hypothesis (BF_{10} = 0.132).

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.707	2.409	1.000	
ICs	0.500	0.293	0.415	0.415	0.504

Note. All models include subject

Table S4.5. Bayesian ANOVA on the PSE.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	0.883	7.556	1.000	
ICs	0.500	0.117	0.132	0.132	1.045

Note. All models include subject

Table S4.6. Bayesian ANOVA on the CoV.

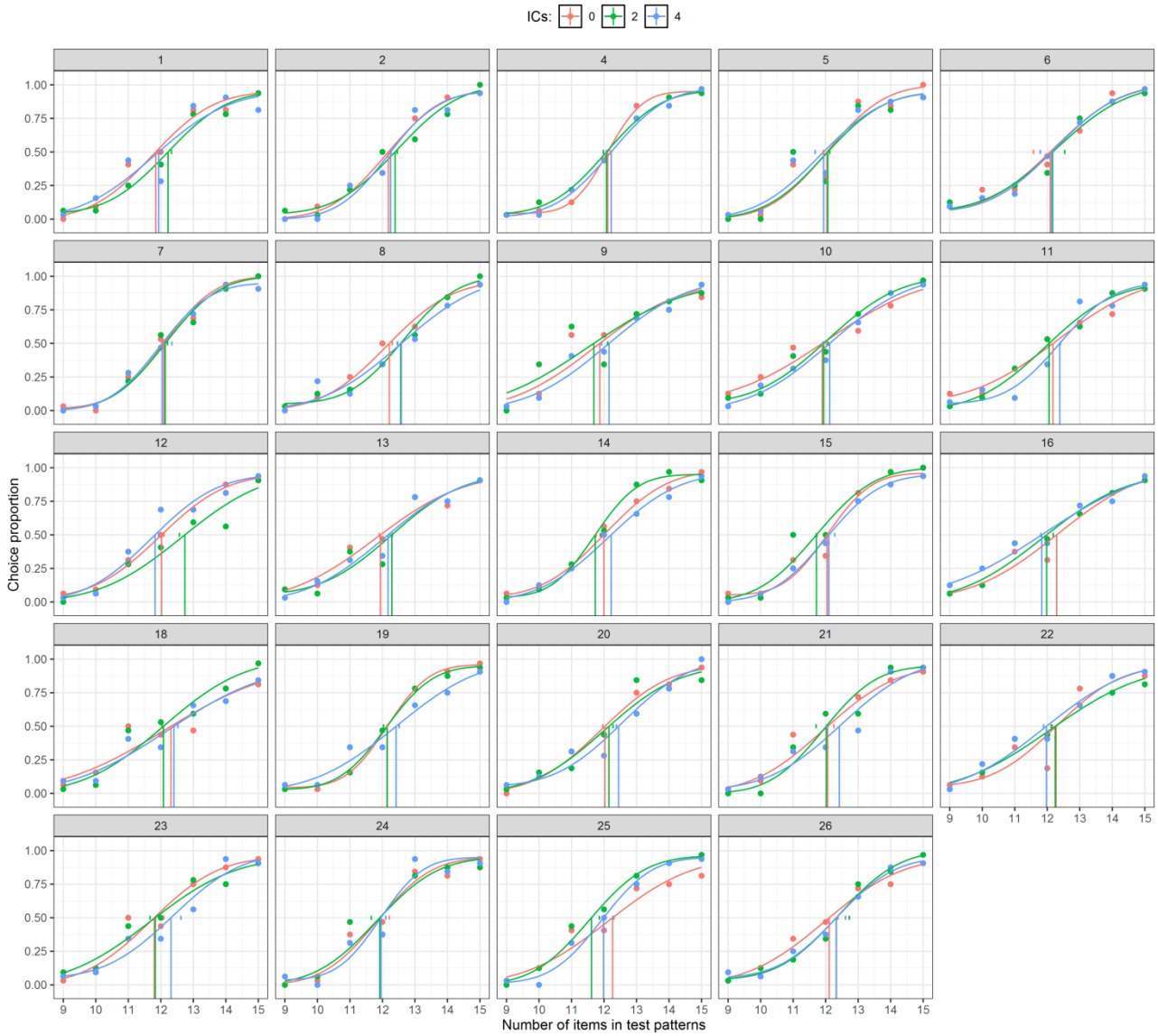


Figure S4.5: Individual psychometric functions for Experiment 3.

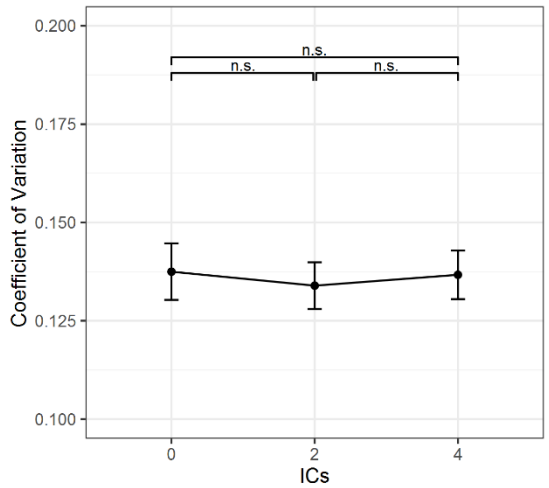


Figure S4.6: Coefficient of Variation as a function of each condition. The error bars represent ± 1 SEM.

Chapter 5

Experiment 1

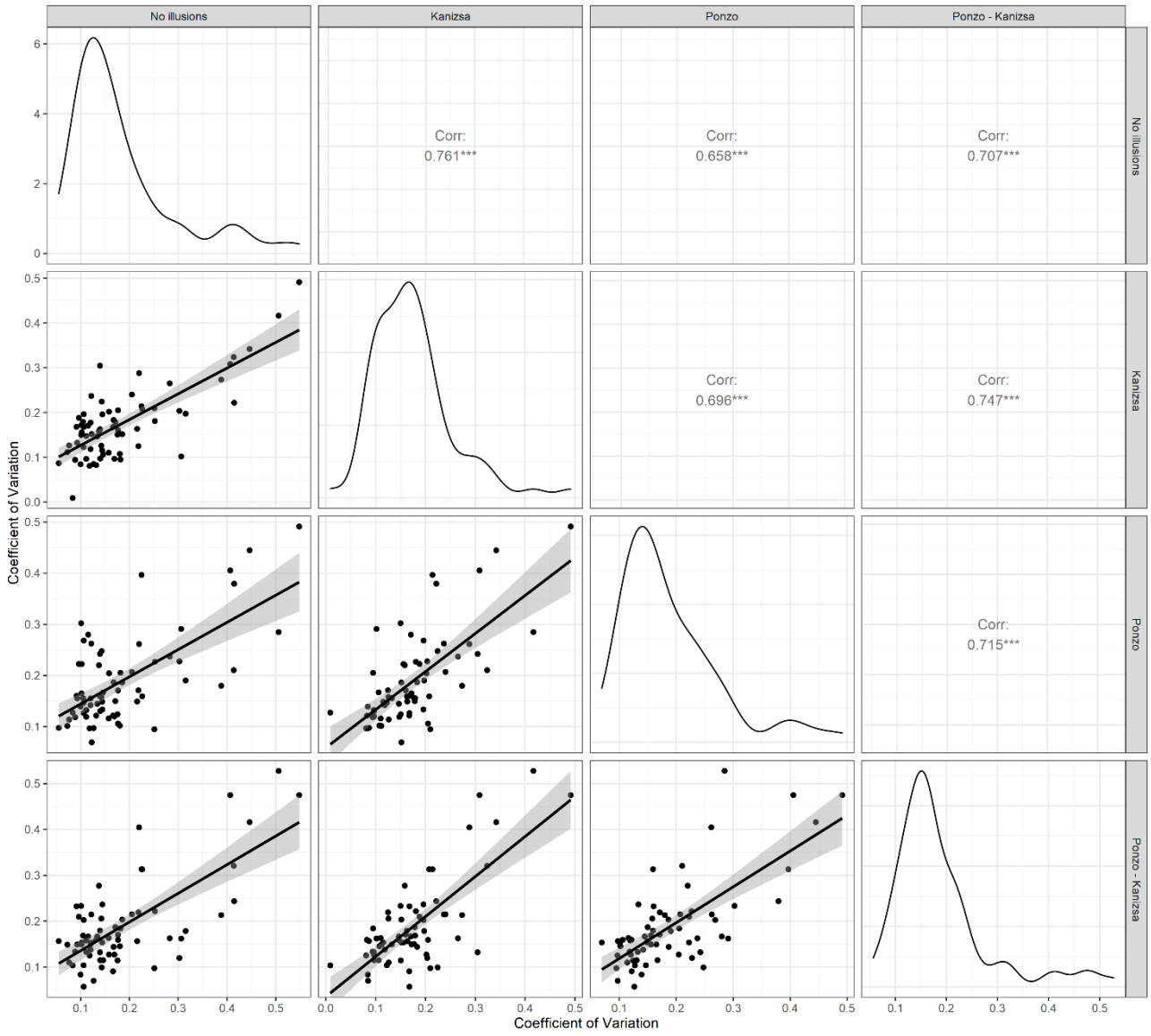


Figure S5.1: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Variable		Kanizsa	Ponzo	No illusions	Ponzo & Kanizsa
1. Kanizsa	n	—			
	Pearson's r	—			
	p-value	—			
2. Ponzo	n	67	—		
	Pearson's r	0.696 ***	—		
	p-value	< .001	—		
3. No illusions	n	67	67	—	
	Pearson's r	0.761 ***	0.658 ***	—	
	p-value	< .001	< .001	—	
4. Ponzo & Kanizsa	n	67	67	67	—
	Pearson's r	0.747 ***	0.715 ***	0.707 ***	—
	p-value	< .001	< .001	< .001	—

* $p < .05$, ** $p < .01$, *** $p < .001$

Table S5.1: CoV correlation values among conditions.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	4.701e -31	1.881e -30	1.000	
Distance	0.200	0.995	843.858	2.117e +30	0.672
Distance + Condition	0.200	0.005	0.019	1.003e +28	1.778
Distance + Condition + Distance ✻ Condition	0.200	4.015e -6	1.606e -5	8.541e +24	1.864
Condition	0.200	2.032e -33	8.128e -33	0.004	0.595

Note. All models include subject

Table S5.2: Bayesian ANOVA on the CoV.

Experiment 2

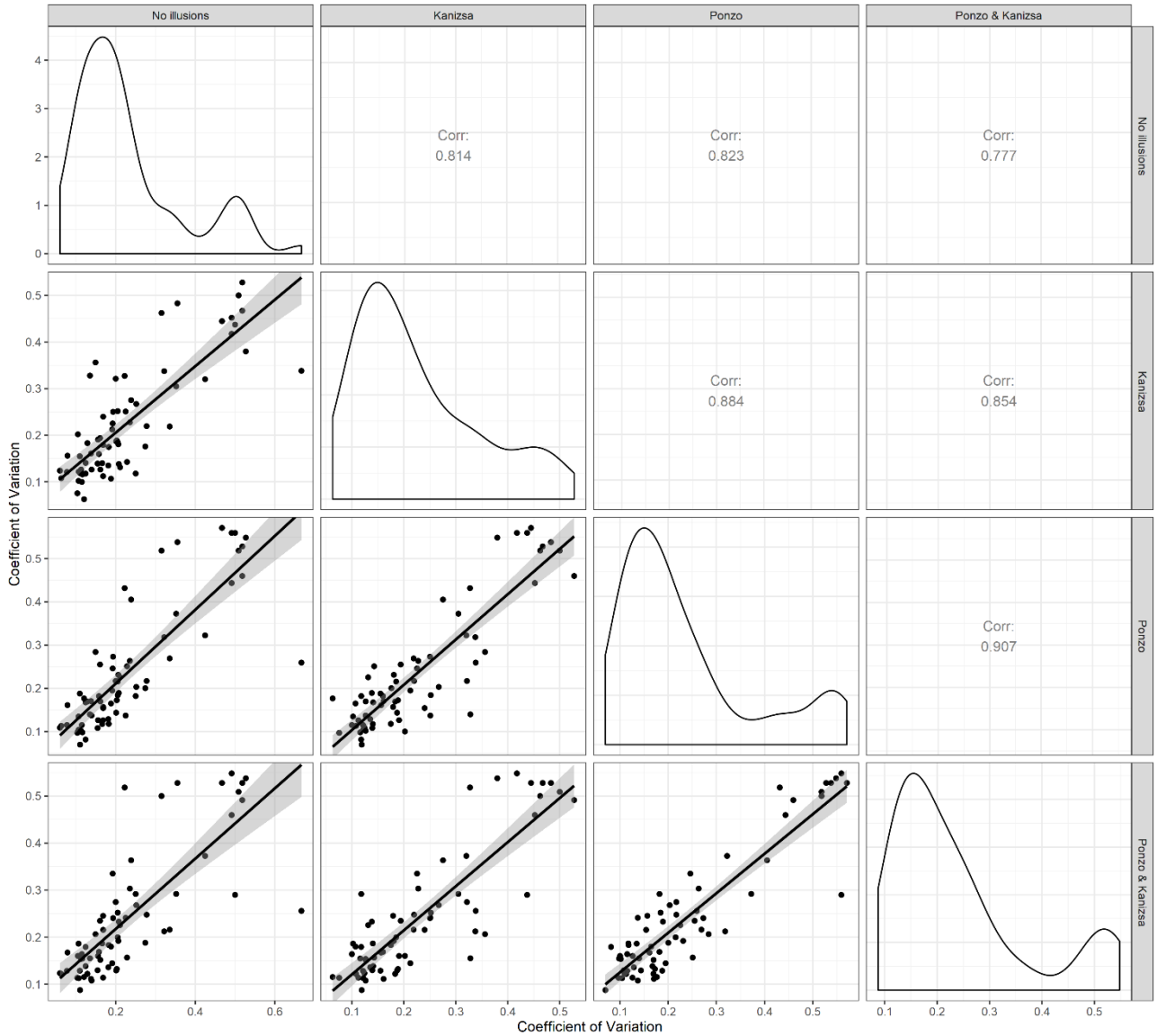


Figure S5.2: Scatterplots showing the correlations among conditions. Shaded regions represent the 95% CI of the correlation line.

Variable		Kanizsa	Ponzo	No illusions	Ponzo & Kanizsa
1. Kanizsa	n	—			
	Pearson's r	—			
	p-value	—			
2. Ponzo	n	68	—		
	Pearson's r	0.884 ***	—		
	p-value	< .001	—		
3. No illusions	n	68	68	—	
	Pearson's r	0.814 ***	0.823 ***	—	
	p-value	< .001	< .001	—	
4. Ponzo & Kanizsa	n	68	68	68	—
	Pearson's r	0.854 ***	0.907 ***	0.777 ***	—
	p-value	< .001	< .001	< .001	—

* $p < .05$, ** $p < .01$, *** $p < .001$

Table S5.3: CoV correlation values among conditions.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	1.479e -10	5.916e -10	1.000	
Distance	0.200	0.996	1030.224	6.735e +9	0.654
Distance + Condition	0.200	0.004	0.016	2.611e +7	1.238
Distance + Condition + Distance ✻ Condition	0.200	5.826e -6	2.331e -5	39391.509	1.148
Condition	0.200	5.443e -13	2.177e -12	0.004	0.575

Note. All models include subject

Table S5.4: Bayesian ANOVA on the CoV.

Chapter 6

Supplementary Materials: Experiment 1

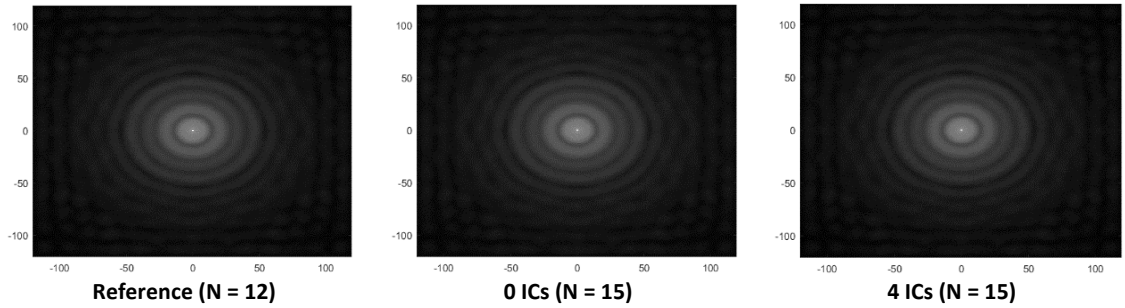


Figure S6.1: Fourier amplitude spectrum displayed as a polar plot for the Reference and the test stimulus with 15 items and 0 or 4 ICs as presented in the Experiment 1. Log-energy is plotted as a function of spatial frequency (distance from the origin; low-to-high) and orientation (angle).

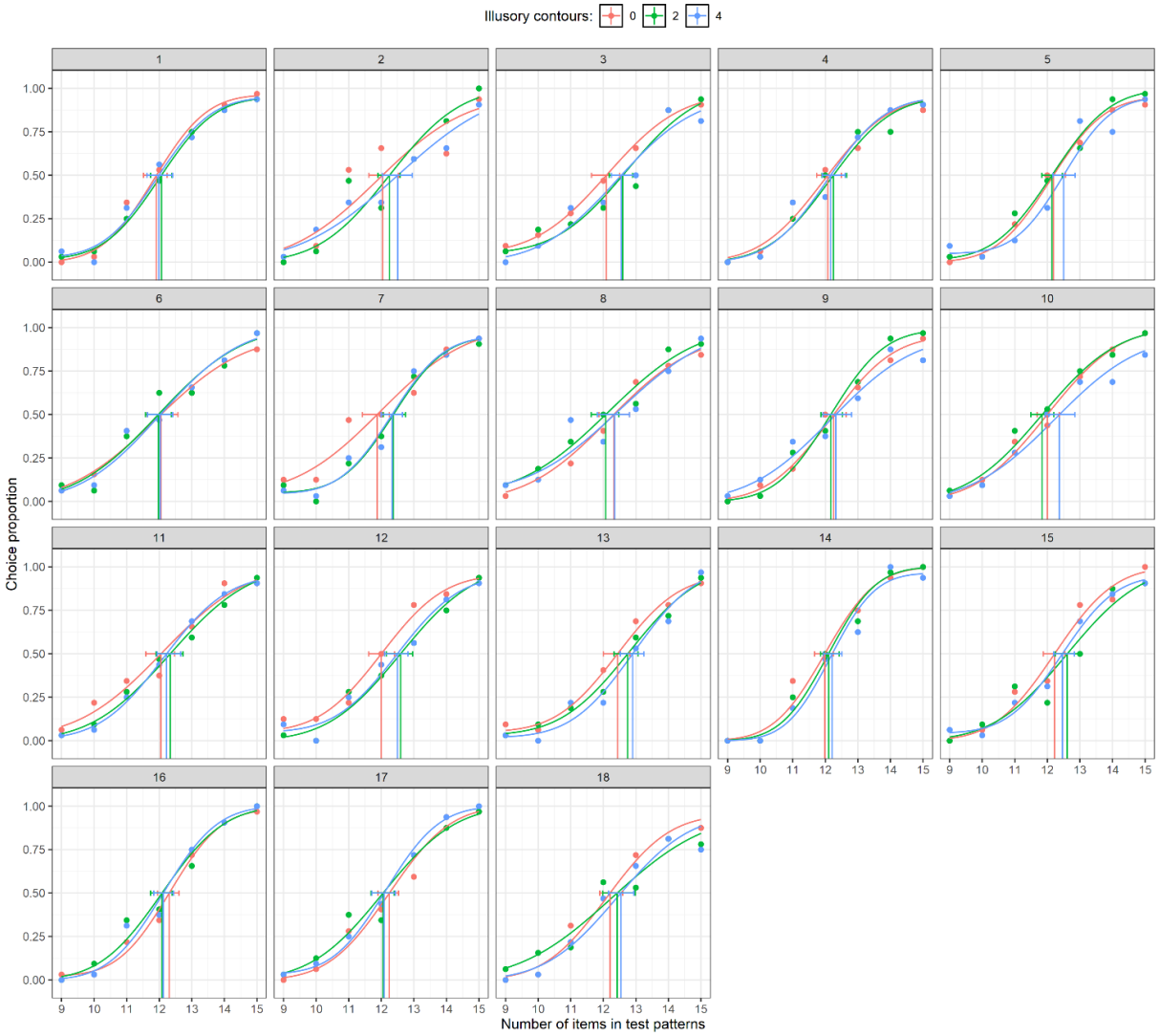


Figure S6.2: Individual psychometric functions for Experiment 1.

Participant	ICs	deviance	p
1	0	3.79	0.98
	2	0.32	1
	4	8.08	0.525
2	0	23.58	0.42
	2	17.68	0.105
	4	4.84	0.755
3	0	0.63	1
	2	7.48	0.825
	4	8.66	0.71
4	0	5.32	0.795
	2	5.78	0.805
	4	5.83	0.915
5	0	1.99	1
	2	3.21	0.98
	4	8.84	0.37
6	0	1.67	0.995
	2	7.96	0.95
	4	3.70	0.915
7	0	4.68	0.815
	2	7.18	0.955
	4	4.29	0.89
8	0	2.45	0.98
	2	2.50	0.975
	4	8.68	0.91
9	0	2.36	1
	2	3.34	0.865
	4	5.06	0.845
10	0	1.26	0.99
	2	2.60	0.94
	4	3.38	0.985
11	0	4.44	0.745
	2	1.17	1
	4	0.47	1
12	0	2.53	0.99
	2	7.30	0.51
	4	8.28	0.995
13	0	1.79	0.97
	2	1.73	0.985
	4	8.76	0.62
14	0	7.89	0.365
	2	6.11	0.675
	4	10.60	0.535
15	0	7.69	0.67
	2	10.75	0.54
	4	2.72	0.975
16	0	1.65	0.99
	2	6.45	0.575
	4	5.39	0.535
17	0	4.80	0.895
	2	4.56	0.99
	4	1.65	1
18	0	5.34	0.835
	2	5.88	0.78
	4	8.90	0.33

Table S6.1: Goodness of fit computation of psychometric functions was performed with *Quickpsy*. All *p* values were > .05, which suggests a good fit of the individual psychometric functions to the data.

Chapter 7

Supplementary Materials: Experiment 1

Visual parameters analysis of the original stimuli

Following the methodology of Gebuis and Reynvoet (2011), we analysed the visual cues using the data generated by their script. A regression analysis was conducted for each visual cue to check the relation between numerical distance and the different visual cues over the original stimuli. The dependent variable for the comparison task was the Weber fraction [(largest number – smallest number) / largest number] and the independent variable was the difference in visual properties of the two stimuli comprising each stimulus pair (Gebuis & Reynvoet, 2011). We reported the R^2 and p values for the following visual cues: area extended (convex hull), total surface (the aggregate surface of all dots in one array), density (area extended/total surface), item size (average diameter of the dots presented in one array), and total circumference (circumference of all dots in one array, taken together).

Results showed that differences in area extended (R^2 = .005, p = .45), density (R^2 = .002, p = .65), total surface (R^2 = .002, p = .62), item size (R^2 = .0006, p = .80) and total circumference (R^2 = .01, p = .28), did not reach the statistical significance and hence could explain only a very trivial (insignificant) part of the variance in numerical distance.

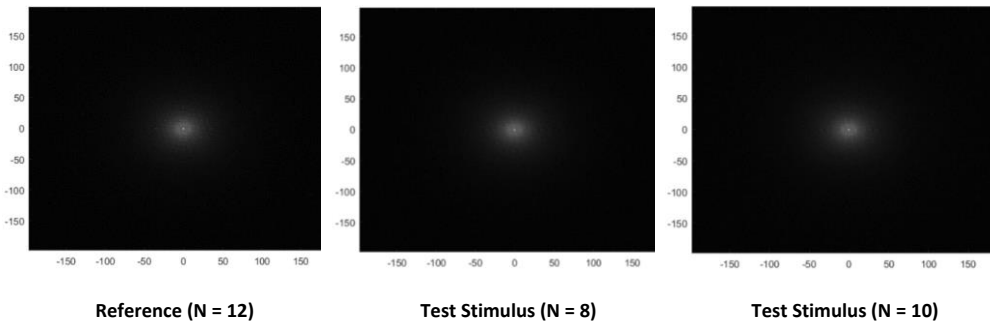


Figure S7.1: Fourier amplitude spectrum displayed as polar plot for the Reference and the test stimulus with 8 items (ratio 0.66) or 10 items (ratio 0.8) as presented in the Experiment 1 (we reported only this information for reasons of space, but a similar pattern was found for the other numerosities/ratios). Log-energy is plotted as a function of spatial frequency (distance from the origin; low-to-high) and orientation (angle).

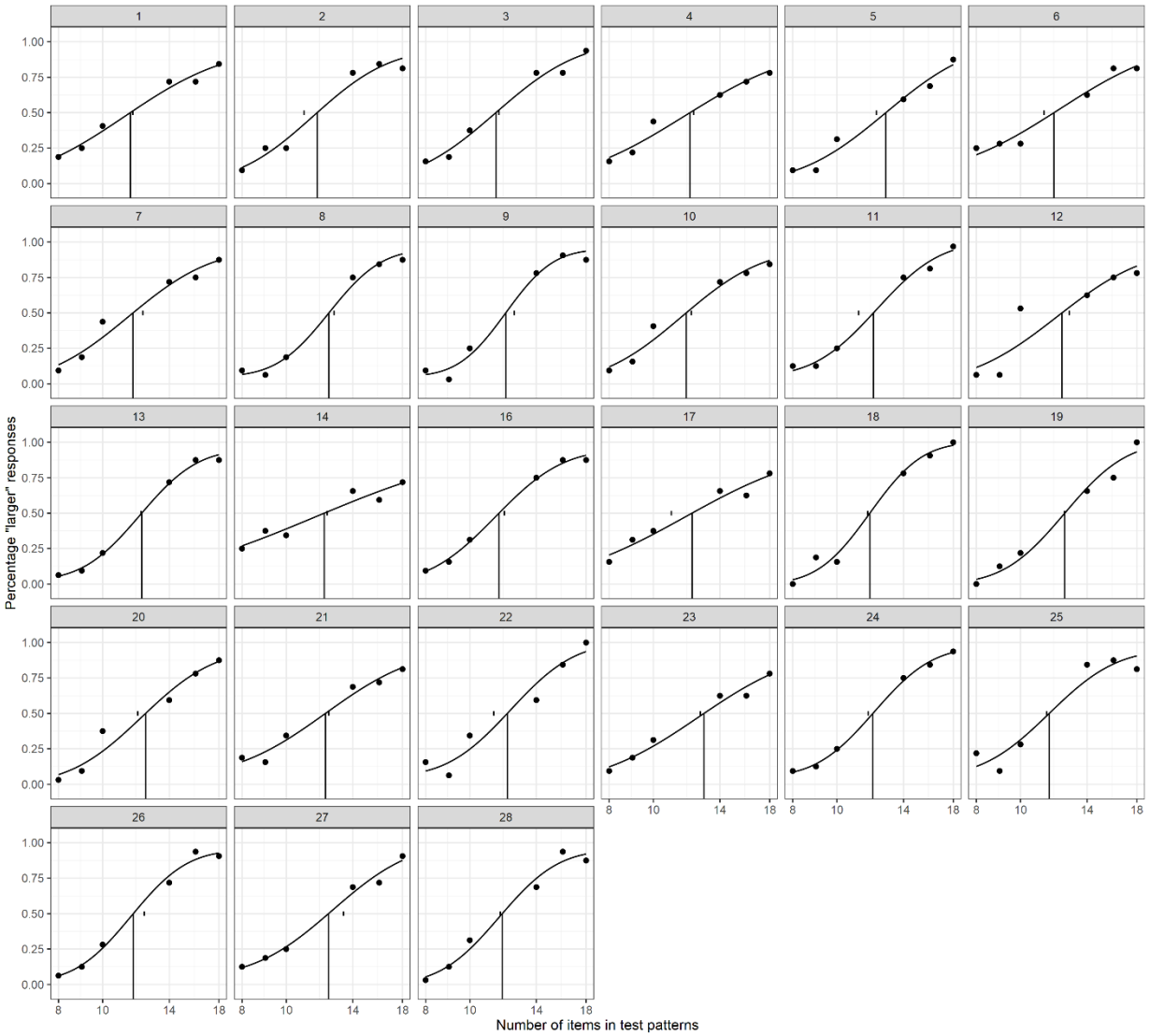


Figure S7.2: Individual psychometric functions for Experiment 1.

Participant	deviance	p
1	1.102	1
2	3.526	0.795
3	2.209	1
4	1.760	0.975
5	2.771	1
6	1.997	0.97
7	3.091	0.845
8	2.173	0.92
9	5.163	0.99
10	2.501	0.905
11	1.943	0.965
12	14.694	0.07
13	0.876	0.98
14	1.867	0.98
16	0.705	0.995
17	1.910	0.98
18	6.743	0.65
19	9.389	0.255
20	5.123	0.835
21	1.960	0.975
22	11.554	0.245
23	1.513	1
24	0.376	1
25	10.189	0.26
26	1.873	1
27	1.364	0.975
28	4.026	0.825

Table S7.1: Goodness of fit of psychometric functions was performed with Quickpsy. All *p* values were > .05, which suggests a good fit of the individual psychometric functions to the data.

Supplementary Materials: Experiment 2

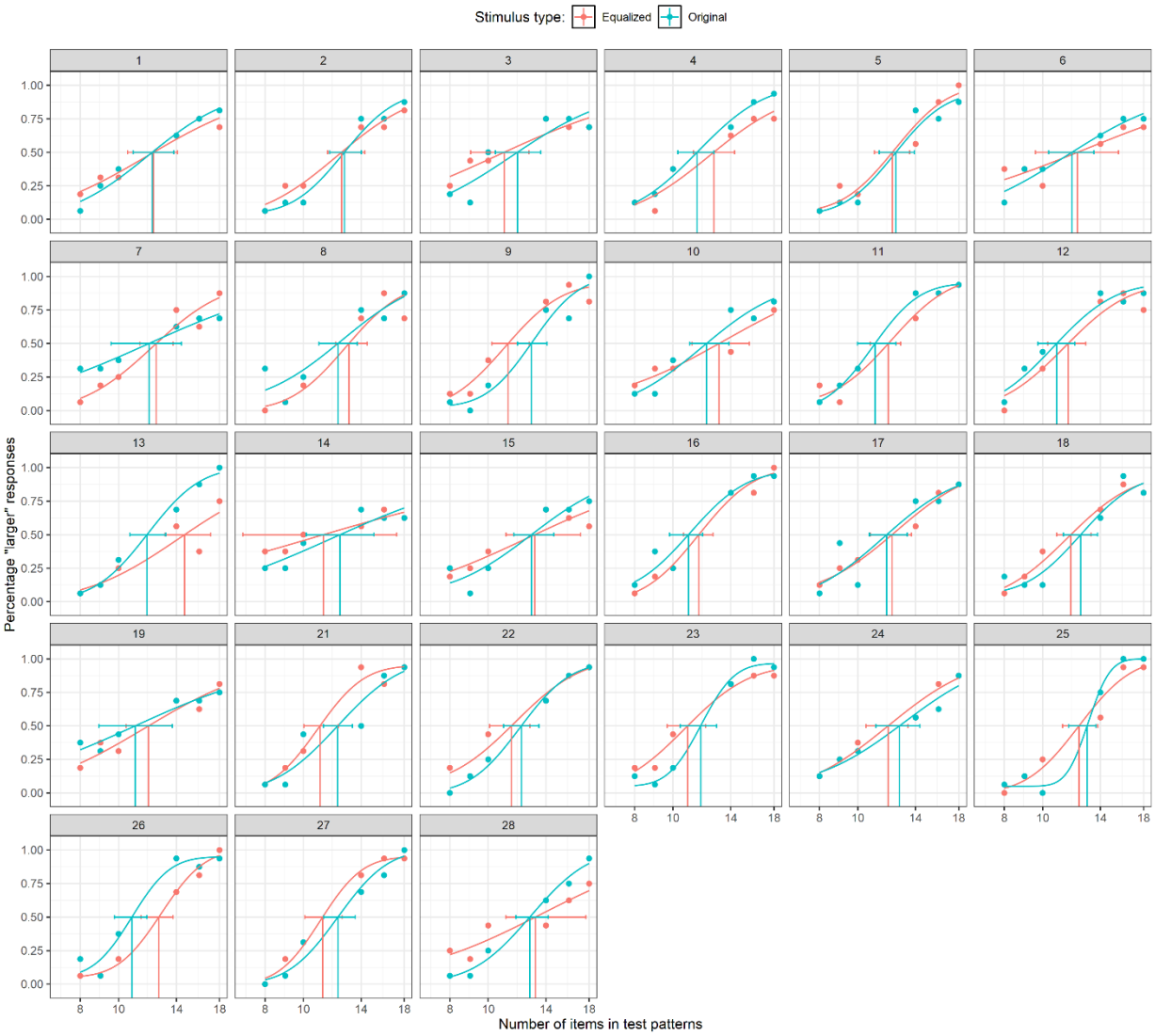


Figure S7.3: Individual psychometric functions for Experiment 2.

Participant	Stimulus type	deviance	p
1	Equalized	0.969	0.98
	Original	1.381	0.99
2	Equalized	1.444	0.975
	Original	1.57	1
3	Equalized	1.912	0.97
	Original	5.604	0.805
4	Equalized	3.295	0.875
	Original	0.516	1
5	Equalized	5.128	0.7
	Original	2.803	0.95
6	Equalized	2.249	0.905
	Original	1.572	0.98
7	Equalized	2.829	0.87
	Original	0.373	1
8	Equalized	6.088	0.89
	Original	7.30	0.36
9	Equalized	3.617	0.75
	Original	7.943	0.66
10	Equalized	1.268	1
	Original	2.466	0.96
11	Equalized	2.945	0.875
	Original	0.544	1
12	Equalized	8.361	0.505
	Original	3.203	0.995
13	Equalized	4.177	0.725
	Original	2.15	0.99
14	Equalized	0.638	0.995
	Original	2.001	0.985
15	Equalized	2.727	0.95
	Original	4.749	0.765
16	Equalized	2.748	0.94
	Original	2.697	0.975
17	Equalized	0.844	0.995
	Original	8.407	0.345
18	Equalized	1.852	0.985
	Original	5.741	0.885
19	Equalized	1.768	0.93
	Original	0.779	0.99
21	Equalized	3.248	0.92
	Original	6.670	0.685
22	Equalized	2.341	0.96
	Original	1.611	1
23	Equalized	1.188	0.995
	Original	3.443	0.99
24	Equalized	1.130	1
	Original	1.468	0.97
25	Equalized	3.988	0.955
	Original	3.868	0.705
26	Equalized	2.139	0.995
	Original	4.945	0.935
27	Equalized	2.018	0.995
	Original	4.770	0.8
28	Equalized	2.424	0.995
	Original	1.29	1

Table S7.2: Goodness of fit of psychometric functions was performed with Quickpsy. All *p* values were > .05, which suggests a good fit of the individual psychometric functions to the data.

Chapter 8

Supplementary Results: Experiment 1

We ran two separate Bayesian repeated-measures ANOVAs (2 x 3) respectively on accuracy and RTs data, with mapping and ratio as independent variables. From an inspection of the table below (Table S8.1) for accuracy, the model that outperforms more the null model is the one with the main effect of ratio only, which received strong evidence in favor of the alternative hypothesis (BF_{10} > 100). We also compared the strength of the Bayes factor for the models that exclude or include the critical interaction term. The evidence *against* including the interaction is roughly a factor of 14, compared to the model with the main factors. This can be obtained as 3.75×10^{44} / $2.75 \times 10^{43} \approx 14$. Thus, the data are almost 14 times more likely under the two main effects model than under the full model (i.e., the one including also the interaction). Finally, the main effect of mapping received substantial support in favor of the null hypothesis (BF_{10} = 0.124). In sum, the Bayesian ANOVA for the accuracy reveals that the data provide very strong support for the main effect of ratio, as well as good evidence against mapping. The data also provide good evidence against including the interaction term.

The analysis of RTs revealed that the model that outperforms more the null model is the one with the main effect of ratio and mapping, and no interaction, which received strong evidence in favor of the alternative hypothesis (BF_{10} > 100) compared to all the other models (Table S8.2). Again, the evidence *against* including the interaction is almost a factor of 10, which can be obtained as 1.46×10^{11} / $1.46 \times 10^{10} \approx 10$. Thus, the data are almost 10 times more likely under the two main effects model than under the full model (i.e., the one including also the interaction).

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	2.949e -46	1.179e -45	1.000	
Ratio	0.200	0.881	29.658	2.988e +45	1.221
Mapping + Ratio	0.200	0.111	0.498	3.755e +44	2.561
Mapping + Ratio + Mapping ✻ Ratio	0.200	0.008	0.033	2.749e +43	4.844
Mapping	0.200	3.650e -47	1.460e -46	0.124	1.932

Note. All models include subject

Table S8.1. Bayesian ANOVA on the Accuracy.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	6.193e -12	2.477e -11	1.000	
Mapping + Ratio	0.200	0.910	40.223	1.469e +11	1.605
Mapping + Ratio + Mapping ✻ Ratio	0.200	0.090	0.398	1.460e +10	14.294
Mapping	0.200	3.181e -6	1.272e -5	513582.968	0.899
Ratio	0.200	2.612e -7	1.045e -6	42168.823	0.750

Note. All models include subject

Table S8.2. Bayesian ANOVA on the RTs.

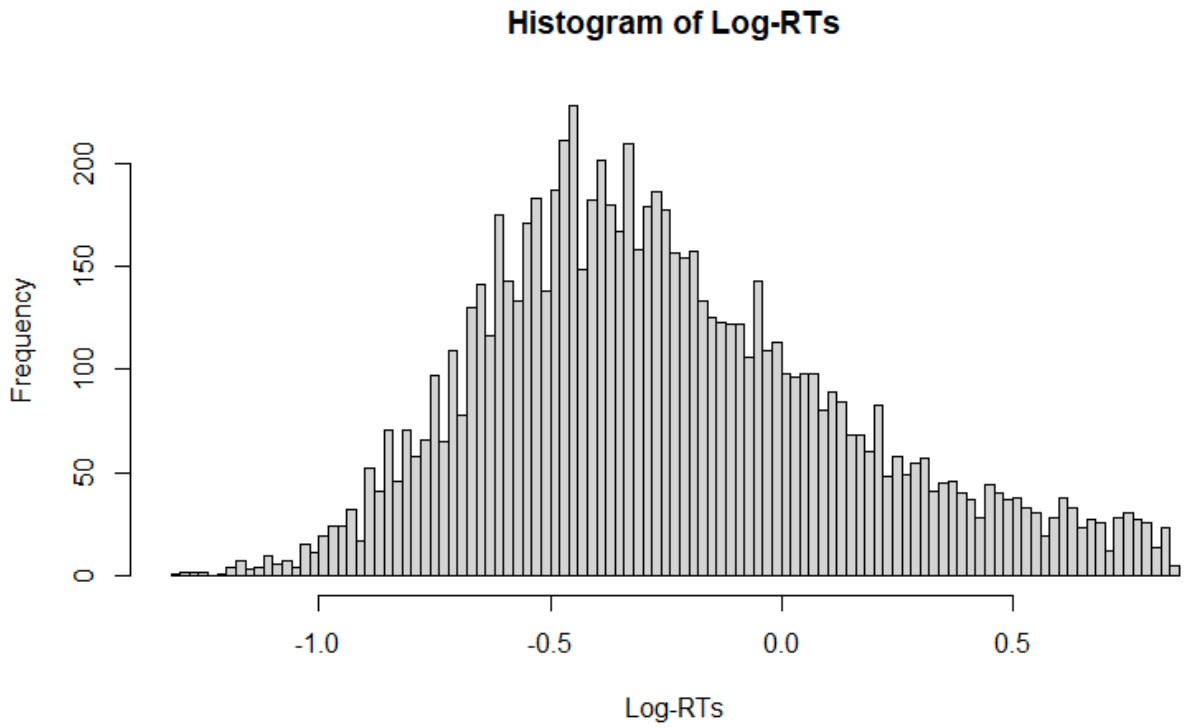


Figure S8.2. log-RTs distribution.

Supplementary Results: Experiment 2

As for the first experiment, we ran two separate Bayesian repeated-measures ANOVAs (2 x 3) respectively on accuracy and RTs data, with mapping and ratio as independent variables. As can be observed in the table below (Table S8.3), the model that outperforms more the null model is the one with the two main effects only, which received strong evidence in favor of the alternative hypothesis (BF_{10} > 100). We also compared the strength of the Bayes factor for the models that exclude or include the critical interaction term. The evidence *against* including the interaction is roughly a factor of 13.5, compared to the model with the main factors. This can be obtained as $3.75 \times 10^{44} / 2.75 \times 10^{43} \approx 14$. Thus, the data are almost 13.5 times more likely under the two main effects model than under the full model (i.e., the one including also the interaction). In sum, the Bayesian ANOVA for the accuracy reveals that the data provide very strong support for the model with the two main effects. The data also provide good evidence against including the interaction term.

The analysis of RTs revealed that the model that outperforms more the null model is the one with the main effect of ratio and mapping, and no interaction, which received strong evidence in favor of the alternative hypothesis (BF_{10} > 100) compared to all the other models (Table S8.4). Again, the evidence *against* including the interaction is almost a factor of 12.5, which can be obtained as $1.027 \times 10^{7} / 821762.006 \approx 12.5$. Thus, the data are almost 12.5 times more likely under the two main effects model than under the full model (i.e., the one including also the interaction).

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	1.950e -58	7.801e -58	1.000	
Mapping + Ratio	0.200	0.674	8.265	3.455e +57	1.649
Ratio	0.200	0.277	1.529	1.418e +57	0.869
Mapping + Ratio + Mapping ✻ Ratio	0.200	0.050	0.208	2.540e +56	1.839
Mapping	0.200	6.812e -59	2.725e -58	0.349	1.680

Note. All models include subject

Table S8.3. Bayesian ANOVA on the Accuracy.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	8.950e -8	3.580e -7	1.000	
Mapping + Ratio	0.200	0.919	45.660	1.027e +7	6.873
Mapping + Ratio + Mapping ✻ Ratio	0.200	0.074	0.318	821762.006	2.337
Ratio	0.200	0.007	0.028	78158.146	1.545
Mapping	0.200	5.488e -6	2.195e -5	61.323	3.687

Note. All models include subject

Table S8.4. Bayesian ANOVA on the RTs.

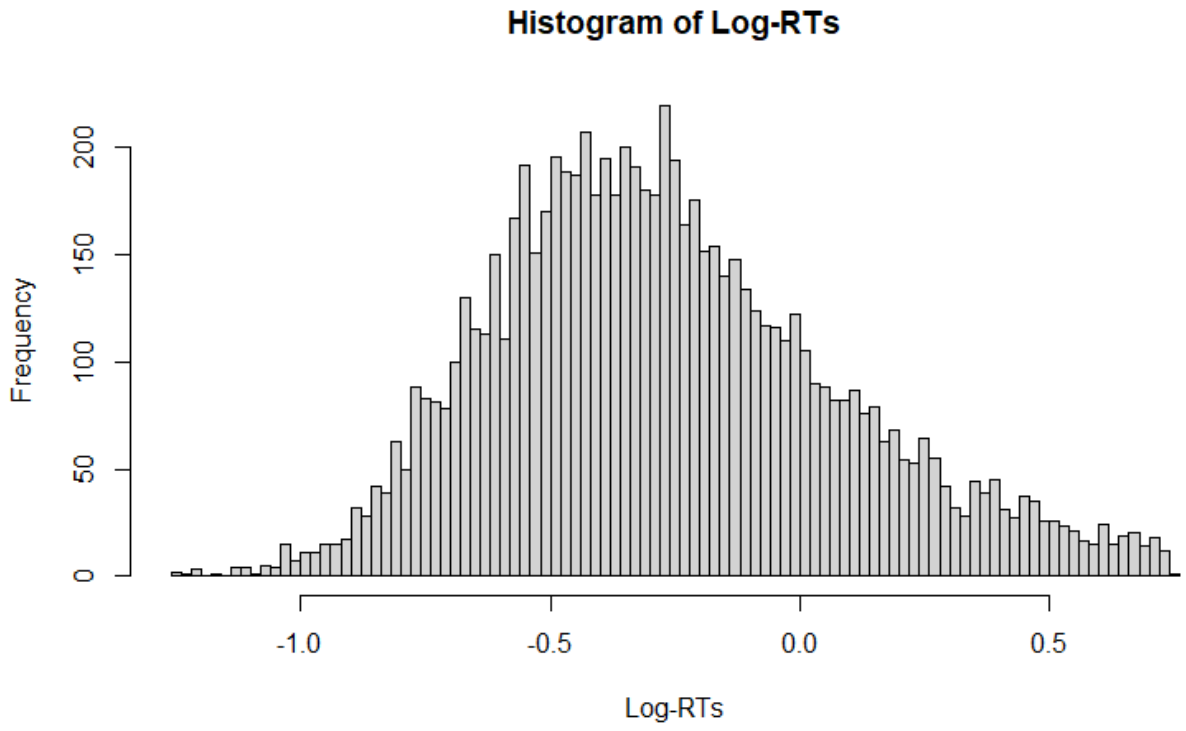


Figure S8.3. log-RTs distribution.

Inter-experiments comparison

As a further control of the analyses reported in the manuscript, we also run an inter-experiments comparison. In particular, we run two three-way mixed ANOVAs (2x2x3) with the Experiment (original vs equalized stimuli) as between-subjects variable, and the mapping (congruent vs incongruent) and the ratio (.66, .75, .8) as within-subjects variables, one over accuracy and one over correct RTs.

Results on accuracy data showed a significant main effect of experiment, $F(1, 102) = 3.94$, $p = .05$, $\eta^2_p = .037$, suggesting a slightly higher accuracy with original stimuli as compared to the equalized ones. Furthermore we found a significant effect of ratio, $F(2, 204) = 393.01$, $\varepsilon = .89$, $p < .001$, $\eta^2_p = .79$, and a significant interaction between ratio and experiment, $F(2, 204) = 5.55$, $\varepsilon = .91$, $p = .004$, $\eta^2_p = .052$, indicating that for higher ratios, discrimination became more challenging (i.e., more prone to errors) with equalized stimuli as compared to the original ones (see Figure S8.4A). The main effect of mapping and all the other interactions were not statistically significant (all $p > .05$). Bayesian analyses confirmed this pattern of results (Table S8.5), since the model that outperforms more the mull model is the one containing the main effects of ratio and experiment

and their interaction, which received strong evidence in favor of the alternative hypothesis (BF_{10} > 100).

Crucially, analysis of correct RTs (4% of data were discarded) revealed only a significant main effect of ratio, $F(2, 204) = 76.63$, $\varepsilon = .88$, $p < .001$, $\eta^2_p = .42$, and a significant main effect of mapping, $F(1, 102) = 15.087$, $p < .001$, $\eta^2_p = .129$ (see Figure S8.4B). The main effect of Experiment and all the other interactions were not statistically significant (all $p > .05$). Also in this case, Bayesian analysis confirmed this pattern of results (Table S8.6), since the model that outperforms more the null model is the one with the main effects of ratio and mapping (and no other interaction), which received strong evidence in favor of the alternative hypothesis (BF_{10} > 100). Furthermore, the main effect of experiment received anecdotal evidence for the null hypothesis (BF_{10} = .34), therefore suggesting that the effect of experiment was 3 times more likely under the null hypothesis (e.g., no difference).

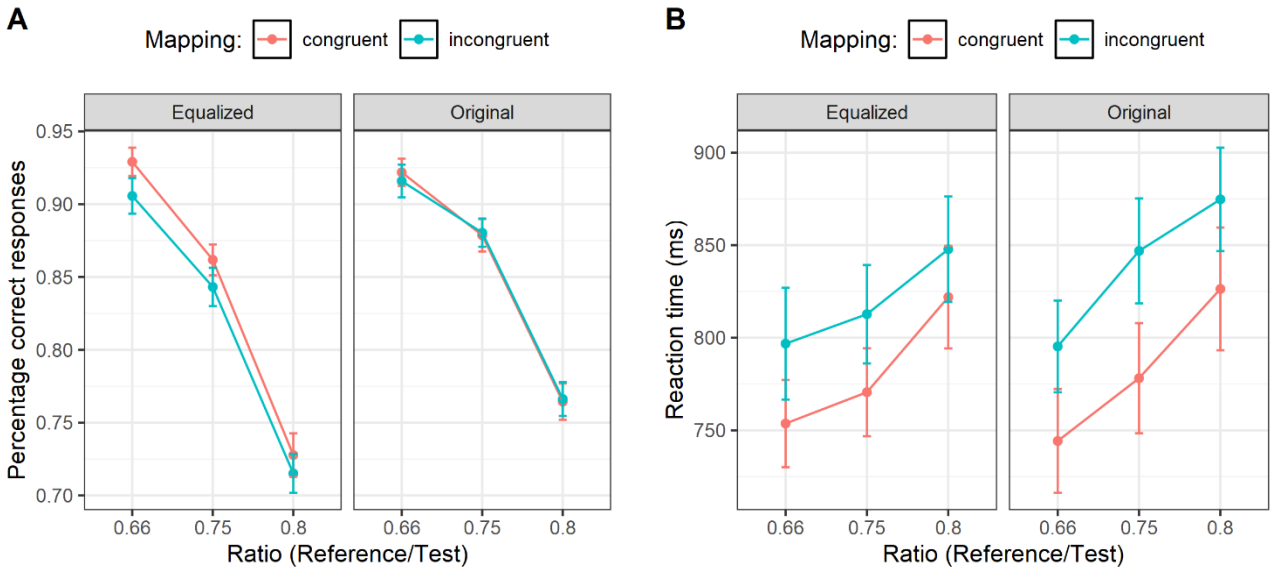


Figure S8.4: A) Percentage of correct responses as a function of the absolute ratio, the mapping condition and the Experiment (i.e., original vs. equalized stimuli). B) Reaction times as a function the absolute ratio, the mapping condition and the Experiment. Bars represent ±1 SEM.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.053	6.197e -106	1.115e -104	1.000	
Ratio + Experiment + Ratio ✻ Experiment	0.053	0.446	14.497	7.199e +104	1.933
Mapping + Ratio + Experiment + Ratio ✻ Experiment	0.053	0.225	5.235	3.636e +104	3.341
Mapping + Ratio + Experiment + Mapping ✻ Experiment + Ratio ✻ Experiment	0.053	0.125	2.561	2.010e +104	5.485
Ratio + Experiment	0.053	0.063	1.212	1.018e +104	5.280
Ratio	0.053	0.052	0.997	8.471e +103	1.565
Mapping + Ratio + Experiment	0.053	0.030	0.563	4.897e +103	3.281
Mapping + Ratio	0.053	0.024	0.447	3.912e +103	1.249
Mapping + Ratio + Experiment + Mapping ✻ Experiment	0.053	0.015	0.272	2.400e +103	3.610
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Ratio ✻ Experiment	0.053	0.010	0.189	1.680e +103	7.978
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Mapping ✻ Experiment + Ratio ✻ Experiment	0.053	0.005	0.091	8.137e +102	3.631
Mapping + Ratio + Experiment + Mapping ✻ Ratio	0.053	0.001	0.025	2.267e +102	6.404
Mapping + Ratio + Mapping ✻ Ratio	0.053	0.001	0.020	1.819e +102	2.187
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Mapping ✻ Experiment	0.053	6.097e -4	0.011	9.839e +101	2.111
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Mapping ✻ Experiment + Ratio ✻ Experiment + Mapping ✻ Ratio ✻ Experiment	0.053	4.080e -4	0.007	6.584e +101	21.206
Experiment	0.053	4.841e -106	8.715e -105	0.781	2.290
Mapping	0.053	1.138e -106	2.048e -105	0.184	7.701
Mapping + Experiment	0.053	8.176e -107	1.472e -105	0.132	2.625
Mapping + Experiment + Mapping ✻ Experiment	0.053	1.777e -107	3.199e -106	0.029	5.202

Note. All models include subject

Table S8.5. Bayesian ANOVA on the Accuracy.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.053	1.304e -20	2.346e -19	1.000	
Mapping + Ratio	0.053	0.516	19.189	3.958e +19	4.047
Mapping + Ratio + Experiment + Mapping ✻ Experiment	0.053	0.205	4.654	1.576e +19	5.490
Mapping + Ratio + Experiment	0.053	0.197	4.429	1.515e +19	4.825
Mapping + Ratio + Mapping ✻ Ratio	0.053	0.027	0.496	2.056e +18	1.842
Mapping + Ratio + Experiment + Ratio ✻ Experiment	0.053	0.016	0.300	1.260e +18	7.308
Mapping + Ratio + Experiment + Mapping ✻ Experiment + Ratio ✻ Experiment	0.053	0.016	0.292	1.226e +18	10.514
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Mapping ✻ Experiment	0.053	0.012	0.210	8.865e +17	9.602
Mapping + Ratio + Experiment + Mapping ✻ Ratio	0.053	0.008	0.152	6.416e +17	7.694
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Mapping ✻ Experiment + Ratio ✻ Experiment	0.053	0.001	0.019	7.960e +16	15.522
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Ratio ✻ Experiment	0.053	8.925e -4	0.016	6.846e +16	10.086
Mapping + Ratio + Experiment + Mapping ✻ Ratio + Mapping ✻ Experiment + Ratio ✻ Experiment + Mapping ✻ Ratio ✻ Experiment	0.053	3.988e -5	7.179e -4	3.059e +15	14.756
Ratio	0.053	3.127e -10	5.628e -9	2.399e +10	0.636
Ratio + Experiment	0.053	1.128e -10	2.030e -9	8.650e +9	4.328
Ratio + Experiment + Ratio ✻ Experiment	0.053	9.070e -12	1.633e -10	6.958e +8	6.787
Mapping	0.053	1.578e -12	2.840e -11	1.210e +8	0.980
Mapping + Experiment	0.053	5.853e -13	1.054e -11	4.490e +7	3.849
Mapping + Experiment + Mapping ✻ Experiment	0.053	4.924e -13	8.864e -12	3.777e +7	4.525
Experiment	0.053	4.460e -21	8.028e -20	0.342	3.447

Note. All models include subject

Table S8.6. Bayesian ANOVA on the RTs.

Individual Regression Coefficients

To further corroborate the regression analyses, we also ran for each experiment an analysis of the individual slopes. Individual data were fitted with a regression model as we did at the group level, this time on each participant (i.e., including the difference between RTs with the right and left hands as dependent variable, and numerosity as predictor). Individual regression coefficients for each subject were thus computed and entered in a one-sample *t*-test to verify whether the individual slopes deviate from zero. Results showed that in both Experiment 1, $t(51) = -3.27$, $p = .001$, $d = .45$, and Experiment 2, $t(51) = -2.29$, $p = .025$, $d = .31$, overall, the regression coefficients were negative and significantly different from zero.